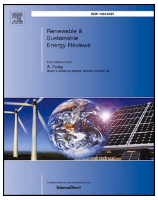
Renewable and Sustainable Energy Reviews 224 (2025) 116095

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser



Advancements in hybrid and ensemble ML models for energy consumption forecasting: results and challenges of their applications

Victor Hugo Rondón-Cordero [a], Lina Montuori [a,*], Manuel Alcázar-Ortega [a], Pierluigi Siano [b,c]

[a] *Institute for Energy Engineering, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*
[b] *Department of Management & Innovation Systems, University of Salerno, Via Giovanni Paolo II, 132, Fisciano, (SA), 84084, Italy*
[c] *Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, 2006, South Africa*

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Hybrid Model
Ensemble method
Energy consumption forecasting
Smart energy management
Smart grid
Microgrid

ABSTRACT

The optimization of energy consumption is crucial for energy management at each level. This research study investigates the different methods that have been used over time to calculate energy consumption using combined machine learning models, specifically hybrid and ensemble models. The first one mentioned combines different machine learning techniques with statistical methodologies that aim to handle non-linear relationships in complex data sets. Some of these representative algorithms are Autoregressive Integrated Moving Average–-Long Short-Term Memory (ARIMA-LSTM), hybrid augmented with Generative Adversarial Network (GAN) and Particle Swarm Optimization–Stacking (PSO-Stacking) which have demonstrated remarkable accuracy in various contexts, including residential, commercial and industrial energy systems. On the other hand, ensemble models that include stacking, boosting and bagging methods have been used to reduce calculation errors and handle large-scale data sets that often have heterogeneous behaviors. The results of this literature review indicate that the selection of an appropriate model that combines different machine learning techniques depends on the nature of the data, the objective and the context of the research, since, as mentioned, hybrid models are more effective in terms of complex, temporal and non-linear data, while ensemble models are more versatile in managing high-dimensional data sets and reducing errors. In addition, the main challenges in this type of work are identified, including computational load and data quality. Given this, it was found that solutions such as the implementation of metaheuristic algorithms and feature selection are available to solve these types of problems.As of this writing, systematic reviews specifically comparing combined machine learning models applied to energy consumption calculation are limited. Therefore, this literature review is a novel starting point for future research that wants to focus on this specific field and needs to solve challenges such as data complexity, time de-pendencies, and computational load, which are oriented to the process of organizing and managing energy consumption. On the other hand, this research offers a perspective that analyzes and contrasts the usefulness and effectiveness of these models in different contexts, as well as identifying advantages, disadvantages, limitations, and opportunities in various areas.

1. Introduction

Energy consumption forecasting is crucial for optimizing resources and ensuring the efficient integration of renewable energy sources (RES) into power grids. Traditional statistical models have been used for decades, but they face limitations in addressing the nonlinear and dynamic nature of energy consumption patterns, especially in contexts influenced by climatic and operational factors [1]. Previous studies demonstrated that energy forecast consumption is essential for microgrids optimal management, for the efficient integration of RES and for the identification of the customer flexibility. Likewise, smart metering [2]and communications technologies [3] are converting traditional energy networks into smart grids capable of collecting large amounts of data, enabling real-time monitoring of consumers' hourly energy consumption [4]. Furthermore, microgrids are another emerging energy system configuration able to change the way the electricity demand is met

---

* Corresponding author.
E-mail address: lmontuori@iie.upv.es (L. Montuori).

https://doi.org/10.1016/j.rser.2025.116095
Received 14 March 2025; Received in revised form 10 July 2025; Accepted 11 July 2025

especially in no-grid-served or isolated areas [5] Research studies present in literature shown the benefits of long short-term memory (LSTM) layer-based artificial neural network model applied to small-scale grids operating in small areas for short-term energy demand prediction. The model resulted in successfully predicted energy consumption with a mean absolute error value of 0.0464 [6]. Despite that, the comparison among different Deep Learning (DL) approaches highlighted the existence of several open issues to still deal with. The high dependence on the efficiency of forecasting methods from historical data, big data, storage and high processing power devices is still a limitation of AI model usage [7]. In response to these limitations, machine learning (ML) techniques have gained popularity.

Speaking specifically about the two types of techniques discussed in this review, hybrid models are characterized by a combination of different heterogeneous methods. For example, a combination of statistical, physical, and mathematical methods that seek to solve problems that cannot be solved individually, but which, when combined, can achieve excellent results, achieving more robust and accurate predictions, taking into account the algorithmic synergy within a unified structure. On the other hand, ensemble models aggregate multiple algorithms to form a collective decision, the objective of which is to improve the stability and performance of predictions, while reduc-ing variance and bias. This is done by leveraging the diversity of the multiple learning models that constitute an ensemble algorithm. Among these, hybrid models and ensemble methods, have shown promising results [8,9]. These approaches have been applied in various contexts, such as smart grids, demonstrating their ability to handle large datasets and capture complex relationships. The implementation of smart grids enabled the collection of a high volume of information. Moreover, the introduction of RES increases the uncertainty in demand forecasting of both microgrids and macrogrids [10]. A combination of Deep Learning (DP), Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) has been highly promoted by tech companies, allowing real-time measurement data and warrantying the cyber security in smart grids Fluctuation of RES generation is one of the major issues also in microgrids [11]. In this sector, ML techniques have shown excellent capabilities to control operation management enhancing the stability, accuracy and optimization of the energy system [12]. In the residential sector, building characterization has always been limited by the large number of inputs required to straighten the gap between as-designed and as-built buildings [13]. ML algorithms have empowered the prediction of residential building consumption by providing measurements every 15 min. In particular, previous studies validate Neural Network-based tools as Least Squares Support Vector Machine as the best predictor for residential electrical consumption However, the literature presents discrepant results. Some studies indicate that hybrid models effectively capture temporal and seasonal patterns best fit for improving the accuracy and stability of the power consumption forecasting by using several artificial neural networks [14,15]. Others highlight that ensemble methods are more suitable for noisy or incomplete data [16]. These statements need to be rigorously supported by deeply analyzing these different techniques, their potential and existing applications and the advantages each of them offers in specific consumption scenarios. In this framework, this research study analyzes about 54 scientific articles and case of study about the use of hybrid models and ensemble methods for energy consumption forecasting. It compared the employed methodologies, model performance, and technical applications across different sectors, such as microgrids, smart buildings and energy systems on a national scale, addressing the topic not only with theoretical studies but also by analyzing practical cases that show how these models work and what optimized results they can obtain. Finally, it provides a comprehensive and unique perspective on these ML emerging techniques identifying opportunities and challenges.

2. Research methodology

The methodology employed to conduct this research study on the latest advancements in hybrid and ensemble ML methods for energy consumption forecasting is shown in Fig. 1 and Table 1.

The primary step consists of the selection of the search database. An indexed database with full-text links is used. The database is chosen to take into account the fields of the research articles, the subject matters, the coverage, and the most relevant publication metrics. The second step is the manuscript search process. The search process has been carried out in 3 stages: pre-screen, in-screen and post-screen process.

During the pre-screen process, the manuscripts have been selected by taking into account the following inclusion criteria.

- Period of publication
- Max Date of publication
- Type of manuscript (Research article, Review, Conference proceedings, theses, and grey literature)
- Topic keywords

The topic keywords will be used to build the search query algorithm that are used to carry out the research.

Before ending the pre-screening, the manuscripts resulting from the search process are ordered and processed as follows.

- Number of citations
- Elimination of the duplicated manuscripts

Secondly, the in-screen process has been done. During this stage, the relevance of the collected manuscripts has been evaluated by taking into account the following inclusion criteria.

- Title affinity with the topic.

The title affinity has been evaluated through a Likert grade (1 – Very irrelevant; 2 – irrelevant; 3 – Somewhat relevant; 4 – Relevant; 5 – Very relevant)

- Abstract interest.

The abstract interest has been evaluated has been evaluated through a Likert grade (1 – Very irrelevant; 2 – irrelevant; 3 – Somewhat relevant; 4 – Relevant; 5 – Very relevant)

Finally, in the post-screen process, the following screening criteria have been taken into account.

- Full-text analysis and Relevance of the results

This parament has been evaluated has been evaluated through a Likert grade (1 – Very irrelevant; 2 – irrelevant; 3 – Somewhat relevant; 4 – Relevant; 5 – Very relevant)

At the end of the screening process, a pool of articles with a total Likert grade upper than 13 points (where 15 points is the maximum value) has been selected.

Once a pool of previous studies has been identified, the most relevant findings for hybrid machine learning models and ensemble methods have been analyzed and discussed in the perspective of their impact on smart grids, microgrids networks and residential and industrial applications.

In the appendices section, the step-by-step process for selecting articles in the different databases is explained, which has been shortened.

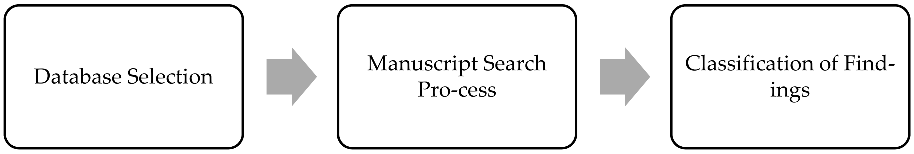


Fig. 1. General workflow of the research methodology.

Table 1
Research methodology.

Stage	Criteria and Activities
1. Database Selection	- Selection of an appropriate indexed database - Coverage of relevant publications - Impact metrics (e.g., citation counts, h-index)
2. Manuscript Search Process	Pre-screen: - Inclusion based on publication period - Maximum publication date - Type of manuscript (research articles, reviews, theses, etc.) - Use of keywords
2.1. Pre-screen Process	- Use of topic keywords to build search queries - Removal of duplicate manuscripts - Sorting by citation count
2.2. In-screen Process	Evaluation criteria: - Title relevance (Likert scale 1–5) - Abstract relevance (Likert scale 1–5)
2.3. Post-screen Process	- Full-text analysis - Relevance of results (Likert scale 1–5) - Final selection of manuscripts with a total score higher than 13 points (maximum 15)
3. Classification of Findings	- Classification based on the impact on smart grids, microgrids, residential, and industrial applications

3. Comparative analysis of hybrid models and ensemble models

3.1. Hybrid Models for Smart Grids

A hybrid machine learning model refers to the combination of several algorithms, which are used together to improve performance and prediction compared to a traditional machine learning model. This is done in order to take advantage of the strengths contained in each model individually. These algorithms are designed to handle complex and non-linear data, which improves the accuracy and capacity in the prediction of the different parameters. These types of models improve robustness against noisy or incomplete data and can better adapt to changing situations, making them more effective in scenarios such as energy systems.

In this research, hybrid models were classified into 2 categories.

- Models based on Neural Networks for Smart Grids
- Models based on Metaheuristic Optimization for Smart Grids

This classification of hybrid models was established according to the prevalence of the consulted literature since most research uses one of these techniques as part of their strategies for the optimization of their results.

3.1.1. Neural network-based models for smart grids

Neural networks are bio-inspired ML algorithm that imitates function and behavior of animal brains. They are versatile and accurate tools to do regression, classificate, forecast of energy patterns dealing with large, high-dimensional data sets.

In the articles cited in this section, it is observed that there are different prominent methodologies such as long short-term memory networks (LSTM), convolutional networks (CNN), bidirectional neural networks (BiGRU) and hybrid combinations that incorporate attention and optimization mechanisms through evolutionary algorithms, which often present better performance than traditional machine learning models, this can be seen in the different comparisons of metrics cited in the studies mentioned in this section.

In this type of hybrid models that base their composition on neural algorithms, they are built based on Deep Learning architectures, which are responsible for understanding and processing different measurements of energy consumption that behave in different ways, for example, methodologies such as CNN-LSTM use convolutional neural networks (CNN) that analyze time series to extract special characteristics from them, such as behavioral patterns that are then loaded into a Long Short-Term Memory (LSTM) network that can model those long-term temporal dependencies, in addition to using this same LSTM model but with attention, they do this with the aim that the network learns to focus on the most relevant parts of the input sequence and thus improve the accuracy of the desired calculation.

Fig. 2 presents a general scheme of neural network-based hybrid models, such as CNN-LSTM or BiGRU, which combine convolutional and recurrent layers to process energy consumption time series.

On the other hand, the articles selected in this section focus exclusively on the application of neural network-based models within smart grids (see Table 4). These studies address key aspects such as energy optimization, load forecasting, fault detection, and grid management. In addition, the articles have also been classified according to the specific field of application, distinguishing between industrial applications (such as factories and production plants), smart buildings, residential (homes and communities) and national and regional consumption. Table 5 shown at the end of this section summarizes this type of classification for all the articles cited (see Table 6).

One of the main articles that talks about the calculation of energy demand in smart grids is [17], in this research the authors developed a model which is a two-level set that combines support vector machines (SVM), convolutional neural networks (CNN-1D) and long short-term memory networks (LSTM). To calculate this consumption, data from the UMass Smart data set was used, which shows the electrical consumption and the different climatic conditions, primarily in residential settings. In addition to using the individual model, it is combined with the shallow neural network technique, which improves the prediction accuracy. Among the authors' conclusions, it can be seen that the hybrid model they developed is more expensive than a conventional model. However, it outperforms traditional approaches with a significant improvement in accuracy, achieving an overall R^2 of 0.956, a Root Mean Square Error (RMSE) of 0.068 (kW), Mean Absolute Error (MAE) of 0.00073 (kW) and a Mean Absolute Percentage Error (MAPE) of 4.2 %, demonstrating the effectiveness of the proposed ensemble model in both short-term and long-term load forecasting.

Another important study is the one conducted by Ref. [18]which focuses on the creation of a hybrid algorithm that improves the prediction of energy consumption in different households. To achieve this goal, a combination of different methodologies is proposed in the hybrid model, mainly Parallel Long Short-Term Memory with Singular Spectrum Analysis (PLSTM with SSA), to correct anomalies during sensor failures and data inconsistencies. This study investigates the prediction of residential energy consumption in smart grids by computing data collected from five households in the UK. The hybrid model gave better results than some reference models, achieving the performance of RMSE $= 0.0043$ kW, MAE $= 0.0028$ kW and MAPE $= 6.67$ % based on data from household 1, giving much faster returns, which can eventually

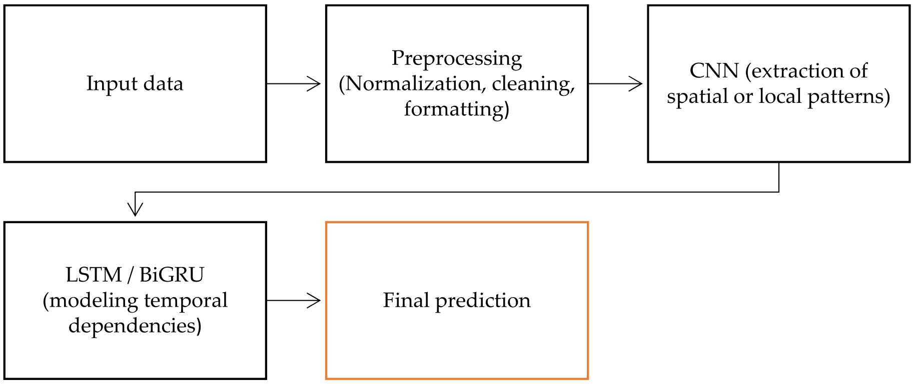


Fig. 2. General architecture of neural network-based hybrid models.

transform into lower costs.

Another model that predicts energy consumption is the one proposed by Ref. [19], this research focused on analyzing different model architectures, among the most important are LSTM, CNN-LSTM, and Sequence-to-Sequence with Attention (Seq2Seq-Attention). The data used for the evaluation of these algorithms were European Network of Transmission System Operators for Electricity (ENTSO-E) and Google Trends, in addition to data processed and standardized with Scikit-learn, which were finally evaluated with metrics such as MSE. The maximum performance was exhibited by the CNN-LSTM model with Google trends having an MSE of 1.045 (RMSE $\approx$ 1022 kW. Converted from reported MSE in MW^2)., while without Google trends, it attained an MSE of 1.076. The LSTM had an MSE of 1.071 with Google Trends and 1.084 without it. The Seq2Seq-Attention had an MSE of 1.155 with an MSE of 1.167 without Google Trends. Extraction of green energy reliable demand forecasting was among the study's outspokenness, pointing to the support placed by Google Trends external data for algorithm utilization. This is most important for its applicability in industrial, economic and residential use, where it provides insight into energy demands in different usage situations.

Similarly, there is a study that analyzes energy consumption in a house in Houston, Texas, where parameters such as the characteristics of the house, meteorological data, and the energy consumption of the house are used to model the algorithms [20]. The models used were Convolutional Long Short-Term Memory (ConvLSTM) and Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) together with two linear models (LSTM and CNN) and their effectiveness was evaluated with the MAE, MAPE and RMSE metrics. The results showed that the ConvLSTM model outper-formed the other algorithms, as they demonstrated superior accuracy with improvements of 4.52 % for 1-day pre-dictions (MAE of 3.69 kWh, RMSE of 5.39 kWh, and MAPE of 18.48 %), 9.59 % for 3-day predictions and 10.53 % for 6-day predictions compared to the other models.

Similarly [21], proposed a hybrid model for the prediction of short-term multi-energy loads but this time applied to industrial areas. The researchers created the model from the combination of CNN and Bidirectional Gated Recurrent Unit (BiGRU). The data used were the energy consumption per hour, in addition to heat and cooling. The objective of the research was to reduce prediction errors and this was achieved, specifically the MAPE decreased by 63.39 % for electricity, 73.03 % for heat and 61.86 % for cooling compared to individual prediction models. For electricity load prediction, the model achieved a MAPE of 2.57 %, RMSE of 652.33 kW, MAE of 462.47 kW, and R^2 of 0.9402.

Continuing with the study of research, it is identified that [22] proposed a two-stage attention-based long-short-term memory neural network model Enhanced Dual Attention Long Short-Term Memory (EDA-LSTM) for energy demand prediction, using daily meteorological and electricity demand data from Pudong, Shanghai (2015–2018). This study focused on residential energy consumption, particularly in the context of Smart Grids. As expected, this combination of models improved energy consumption prediction over individual algorithms with a (MAPE) of 4.57 % and additional metrics, including a MAE of 3111.55 and RMSE of 4015.04 for a 100-day prediction period.

Speaking specifically about industrial energy [23], make a study where they develop a hybrid model that combines Adaptive Neurofuzzy Inference systems (ANFIS) with gene expression programming (GEP), using data such as monthly historical data on electricity consumption, along with environmental factors such as temperature and humidity, and temporal variables such as month and year. This study addresses the plans for long-term energy consumption in industries; in this case, the authors examined the food processing sector in Uganda. The authors concluded that their hybrid GEP-ANFIS model outperformed individual algorithms with respect to their prediction, achieving a MAPE of 0.1934 %, an MAE of 0.0005, and an RMSE of 0.0007. The model further indicated very good performance with an R^2 coefficient of 0.9841. This research is directly applicable to Smart Grids, with industrial load forecasting offering the most accurate planning approach towards the optimization of energy distribution in contemporary power networks.

Similar to the previous study [24],we propose a hybrid algorithm that combines Improved Convolutional Neural Networks (Improved-CNN), Bi-Directional Long Short-Term Memory (Bi-LSTM), Graph Neural Networks (GNN), Transformer Models, and fusion layer in order to obtain more accurate and less error-prone calculations. Compared to other simpler approaches, this hybrid algorithm shows excellent results with RMSE: 5.7532 Wh, MAE: 6.7532 Wh, MAPE: 3.5001 % and R^2: 0.9701. The study utilizes historical energy consumption data collected from smart buildings, focusing on improving energy load predictions within the context of Smart Grids.

Analyzing other research, it is interesting to mention the work by, [25].The authors of this work propose to build a hybrid model based on variational modal analysis (VMD) with bi-directional recurrent units (Bi-GRU), with the aim of calculating energy consumption in smart homes. It is important to mention that the authors used a VMD algorithm to decompose the data that were presented as time series, into intrinsic modal functions (IMF). To build the algorithm, the variables that were used were: energy consumption as the dependent variable, meteorological variables, temperature, humidity, dew point and time of day as independent variables. Among the main results of the model, different metrics were found (MSE of 0.0038 kW, RMSE: 0.0616 kW MAE of 0.046 kW, MAPE of 0.11 % and an R^2 of 0.98) which indicate that the algorithm had excellent precision in calculating and optimizing energy consumption.

Continuing with the analysis of research, another important work is the one developed by Ref. [26].Here the authors propose an algorithm based on parallel long-term memory neural networks (Parallel LSTM, PLSTM) and singular spectrum analysis (SSA), which is a hybrid model that allows calculating energy consumption, which was measured with sensors in residential smart grids, where, specifically, it contains electricity data in household appliances in five homes in the United Kingdom. Some of the values that can be observed in the simulation results are the following: MAE ranging from 0.0020 to 0.0028, MAPE between 4.06 % and 10.55 %, and RMSE from 0.0037 to 0.005. For the best-performing household, the Singular Spectrum Analysis–Parallel Long Short-Term Memory (SSA-PLSTM) model achieved a MAE of 0.0028 kW, RMSE of 0.0043 kW, and a MAPE of 6.67 %.This indicates that the model has a high precision capacity with respect to other individual algorithms and conventional statistical techniques.

Using a slightly different methodology than the one discussed in this section [27], create a hybrid model combining Gated Recurrent Unit (GRU), Multi-Modal Attention (MMattention), and Light Gradient Boosting Machine (LightGBM), integrated with Prophet-based feature extraction. Like the algorithms described earlier in this review, the main goal of these investigations is to make the model more accurate and optimal than individual algorithms, this is achieved by taking advantage of the different advantages of each model: Prophet to decompose time series data into components such as trend, seasonality and residuals; GRU-MMattention to capture temporal dependencies; and LightGBM for parameter optimization. The results found as expected were successful, a 8.66 % reduction in relative error compared to GRU alone and a 2.02 % improvement over LightGBM could be observed. Metrics such as MAPE, MAE and RMSE further validated their performance, showing optimal values. The study utilizes historical energy consumption data from regional power systems (PJM) to enhance prediction accuracy within Smart Grids. The final model achieved a MAPE of 1.69 %, MAE of 157.07 MW, and RMSE of 207.68 MW.

Similarly [28], uses the advantages of Prophet and LSTM models, to create a hybrid algorithm that specifically works on maintaining seasonal patterns while capturing intricate temporal relationships in descaled data, and in this case, calculating energy consumption. The study analyzed monthly energy consumption data from seven countries: Canada, France, Italy, Japan, Brazil, Mexico, and Turkey, from 2006 to 2017, focusing on national energy consumption with a potential application in traditional power grids and Smart Grids. Similar to the various hybrid models mentioned, this algorithm provides better results compared to individual models such as independent neural networks and other statistical techniques, achieving MAE values in the range of 845.159–8880.816 and RMSE values from 1051.658 to 4119.490, depending on the nation being studied. This adds to an ever-growing number of articles indicating how hybrid models based on the decomposition approach do a better job than some of the other independently working algorithms. The article does not report MAPE or R^2 values for the evaluated models.

Similarly, the paper proposed by Ref. [29] proposes an algorithm for forecasting electricity demand by coupling empirical modal analysis (EMD) networks with long short-term memory (LSTM) networks. This hybrid system is used to perform time series prediction tasks in the energy domain with higher accuracy. This study mainly focuses on regional electricity consumption, and it can find application in traditional power grids and Smart Grids because of their critical importance to grid stability and energy management. EMD helps to extract useful patterns from the data, while LSTM networks naturally model the temporal behavior of virtually every mode. The first step of this method is to apply EMD to simplify the analysis of a complicated time series of electricity demand data into intrinsic modal functions (IMFs). The IMF reflects distinctive oscillatory modes within the data, thereby isolating certain frequency components. The LSTM networks model and produce forecasts for each IMF separately, thus taking full advantage of advances in long-term dependencies in sequential data. The final forecast is then obtained by fusing the individual forecasts of each IMF into the original signal. Empirical tests carried out on statewide electricity demand data for California ISO encompassing regional energy usage reported RMSE of 278.76, MAPE of 9.52 %, R-squared of 0.9221 for 24-h forecasts, and were observed to outperform traditional forecasting methods. The EMD-LSTM hybrid substantially outperforms more classic forecasting techniques, that is, single-mode LSTMs and classical statistical methods. This synergistic work improves forecasting accuracy and enables to determine the complexities inherent in electrical load time series during capture, which is extremely effective.Another relevant research that explores this type of models is the one carried out by Ref. [30], where the authors present a hybrid algorithm for the prediction of energy consumption (Fuzzy Recurrent Neural Network (Fuzzy-RNN)), which combines fuzzy logic and recurrent neural networks (RNN). To build the model, energy demand data were used along with different climatic variables such as dry bulb temperature and dew point temperature, which presented non-linear relationships between them. Hourly consumption was measured from January 2013 to December 2014, taking into account the different patterns presented on weekdays, weekends and holidays, in order to observe their fluctuations depending on the circumstances of the day. The authors conclude that the predictions of this algorithm are significantly better than other conventional models (RNN, ANN, SVM or GRNN: 1.76 % MAPE). These were achieved thanks to the combination of models and the use of fuzzy logic.

Another interesting hybrid model to analyze is the one proposed by Ref. [31], which combines long-short-term memory (LSTM) networks with an improved sine-cosine optimization algorithm (ISCOA). Like the previous models presented, the main objective of this algorithm is to calculate and optimize energy consumption. Different variables were used to build it, among the main ones were: climatic variables such as temperature and humidity, historical data on energy consumption, occupancy patterns and temporal characteristics, which were extracted from measurements in smart buildings. According to the MAE, MAPE and RMSE metrics, it was concluded that the model was successful compared to other individual algorithms, since it achieved precision improvements of 15–25 %. In the short-term prediction scenario, the ISCOA-LSTM model achieved a MAE of 0.0610, RMSE of 0.0861, and MAPE of 5.2961 %.

On the other hand, the one proposed by Ref. [32] proposed a hybrid model designed to predict energy consumption, combining empirical modal investigation (EMD), least squares support vector regression (LSSVR) and quadratic exponential smoothing (QES). The context of the study is developed in the cement grinding process, where production conditions, control methods, material composition and fineness are taken into account for the calculation of energy consumption. There was considerable empirical evidence to assert that the hybrid EMD-LSSVR&QES model outperformed the traditional models showing RMSE of 10.81 kWh and MAPE of 0.13 % for one-step-ahead forecasts. The model's effectiveness in explaining such variations in energy demand was also attested to by a trend direction accuracy (Dstat) of 85.34 %. The study focuses on industrial energy consumption and is particularly relevant for optimizing energy usage in large-scale manufacturing processes within industrial grids. Like most hybrid models, this algorithm allowed to improve the results compared to individual models, in addition to the optimization of time and other variables.

Similarly [33],developed a hybrid model that combines an adaptive differential evolution algorithm (ADE) with back propagation neural networks (BPNN), for the calculation of energy consumption such as EEC (electrical energy consumption) and TEC (total energy consumption). To create the algorithm, the authors used variables such as GDP, population, imports and exports. The results of the research showed that the ADE-BPNN model outperformed individual energy calculation methods since a 44.96 % reduction in MAPE was achieved. In addition, the study highlighted the main influence of GDP and population on energy consumption, contrary to exports, which were identified as a less significant factor than the other variables studied. The best performance was obtained with a MAPE of 1.639 %, MAE of 2.0141 GWh, and RMSE of 2.9591 GWh.

A final algorithm analyzed in the compendium of hybrid models was the one proposed by Ref. [34], where the authors present a model applied to solar energy consumption in Denmark, using variables such as wind speed, temperature and solar production for training the algorithm. Basically, the model is a combination of CNN with LSTM, which achieved an RMSE of 7.5 % for a 24-h forecast, 7.7 % for a 48-h forecast and 7.9 % for a 72-h forecast, compared to traditional models, such as Random Forest and Support Vector Machine. Finally, the authors can conclude that this combination of CNN with LSTM is quite effective as a hybrid model compared to other individual models.

The following table summarizes findings about Neural Network-Based Models used in energy consumption forecasting.

Table 2 presents different hybrid models used for energy consumption prediction, grouped by similar algorithms based on the references included in this section. The models are classified into categories such as LSTM variants (LSTM, CNN-LSTM, BiGRU, PLSTM), hybrid models with statistical techniques (SSA, VMD, EMD, QES), neuro-fuzzy and evolutionary models (ANFIS, GEP, ADE), attention-based hybrid models (EDA-LSTM, Transformer, GRU-MMattention), and convolutional and Sequential hybrid models (CNN, Improved-CNN). Additionally, the table highlights the specific applications of each model, indicating whether they are focused on residential, industrial or regional energy consumption.On the other hand, Fig. 2 shows the MAPE (Mean Absolute Percentage Error) value, which is a value that predicts how far the model's prediction value is from the actual value.

Fig. 3 details the MAPE value in the different machine learning models grouped by application scenarios: residential, smart buildings, industrial, national, and regional. The first thing that stands out is that the models generally have a MAPE of less than 5 %, and specifically the models applied to industries have the lowest values. This could be because in industrial processes, energy consumption may follow repetitive patterns, in automated processes with less human variability, and because measurement systems are more precise, compared to, for example, residential complexes where there is heterogeneity in human energy consumption. It is worth noting that two investigations did not present a real MAPE value, therefore, they do not present a numerical relationship in Fig. 1. Despite the above, the model with the lowest MAPE was the VMD + BiGRU [25] applied to smart homes, with a MAPE of just 0.11 %. VMD (Variational Mode Decomposition) is used to decompose the power consumption series into several simpler modes, eliminating noise and abrupt fluctuations. In addition, the bidirectional version (BiGRU) reads information forward and backward in time, which better captures complex temporal relationships.

Although it was mentioned that the best performer was the industrial field, the use of this type of algorithm and the particularity of being a smart home, with better-controlled conditions for data measurement, could have influenced the lower MAPE compared to the others.

Table 2
Neural network-based models.

Studied Models	Main Characteristics	Applications & References
LSTM Variants	It capture temporal dependencies.	[17] - Residential [18], - Residential [19], - Regional [20], - Residential [21], - Industrial [22], - Residential [29], - Regional
Hybrid models with statistical techniques	It integrates statistical techniques like SSA, VMD, EMD, and QES.	[18] - Residential [25], - Residential [29], - Regional [32], - Industrial (Cement)
Neuro-Fuzzy and Evolutionary Models	It combination is to handle non-linear relationships.	[23] - Industrial (Food Processing) [31], - Smart Buildings [33], - National
Attention-Based Hybrid Models	It uses to improve prediction accuracy and reduce errors.	[24] - Smart Buildings [27], - Regional [28], - National [32], - Industrial
Convolutional and Sequential Hybrid Models	It is for better temporal modelling.	[24] - Smart Buildings [34], - Regional energy consumption

3.1.2. *Metaheuristic optimization-based models for smart grids*

In this type of hybrid algorithms, their construction is based on models such as Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO) or Jellyfish Search, which are combined with machine learning or Deep Learning in order to optimize internal parameters, searching for different configurations of hyperparameters that minimize errors in the prediction of energy consumption, that is, they behave as optimization tools for machine learning models. Some of these algorithms are described below:

Particle Swarm Optimization (PSO): This technique is scientifically based on the behavior of swarms. Each particle studies the space of different solutions and then adjusts its position based on previous experience.

GWO (Grey Wolf Optimizer): This algorithm is inspired by the social hierarchy of wolves, whose objective is for the best solutions to guide others in finding new ones.

Jellyfish Search: This algorithm is based on the feeding behavior of jellyfish. It combines global exploration based on random movements with local exploration, the goal of which is to find the most optimal solutions.

This section describes all the hybrid models used to calculate energy consumption using these described methodologies, finally the cited articles were further classified into four categories: residential, smart buildings, industrial, and regional consumption, and this summary is presented in Table 3.

Fig. 4 shows the general structure of hybrid models where metaheuristics optimize machine learning configurations for better prediction performance.

The first study to analyze is the one carried out by Ref. [35], who built a hybrid model that is capable of predicting the energy consumption of buildings, using a model called PSO-Stacking, which integrates particle swarm optimization (PSO) with machine learning algorithms. The main objective of creating the model was to improve the forecasting accuracy while optimizing training time. The three algorithms that supported the feature selection stage were Recursive Feature Elimination (RFE), Random Forest (RF), and Variance Threshold (VT). The variables that were used to train the model were relative humidity, dry bulb temperature, atmospheric pressure, solar radiation, and others. Like other hybrid models, this algorithm outperformed different individual models, which can be validated with the improvement of some symmetrical variables such as MSE of 242.42, RMSE of 15.56, MAE of 10.80, R^2 of 0.97, and MAPE of 5.56 %. The study was successful because, in addition to improving accuracy, training times were significantly reduced. The model was applied specifically to campus buildings using data collected from a university building in Cambridge, Massachusetts.

On the other hand, the study carried out by Ref. [36]developed an algorithm that mixed different metaheuristic techniques, starting from linear and non-linear models which used algorithms such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and Least Squares Support Vector Regression (LSSVR) that were finally optimized by the Jellyfish Search (JS) algorithm, a method that demonstrated superior performance in parameter tuning compared to other metaheuristic techniques. The data used was a set of energy time series parameters, which gave rise to a successful model in terms of accuracy and computational efficiency. It should be noted that this model needed fewer input variables (4 inputs) compared to other models in the literature, which generally used between 7 and 32 variables. The accuracy of the JS-LSSVR (SARIMA, LSSVR) model was demonstrated with an RMSE of 12.84, MAE of 164.86, and R^2 of 0.966, and MAPE of 2.98 % when tested on regional electricity consumption datasets, confirming its superior performance in regional energy supply planning applications.

Another relevant research was carried out by Ref. [37]where a medium-term prediction model was developed for the consumption of electric energy specifically in oil pipelines. The hybrid model uses support (SVM) and genetic algorithms (GA) where SVM is used for the

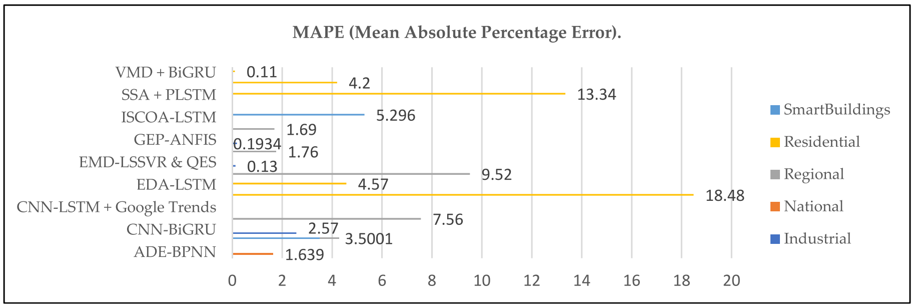


Fig. 3. Mape (mean absolute percentage error).

Table 3
Metaheuristic optimization-based models.

Studied Models	Main Characteristics	Applications & References
PSO-based Models	It combines PSO with various machine learning algorithms like SVM, GRNN, and MLR for optimization.	Smart Buildings [35], National Energy Consumption [40], Industrial (Copper Foil Production) [42]
Genetic Algorithm-based Models	This algorithm is frequently used to optimize different parameters in all types of models, including hybrid ones.	Industrial (Oil Pipelines) [37], Industrial (Oil Pipelines) [39]
Symbiotic and Nature-inspired Models	It is inspired by parameters that exist in real life such as Jellyfish Search, SOS, and GWDO	Regional Energy Consumption [36], Residential [41], Regional Energy Consumption [38]

Table 4
Ensemble Model.

Studied Models	Main Characteristics	Applications & References
Stacking-based Models	It combines base models like Extra Trees Regressor, AdaBoost, Random Forest Regressor, and XGBoost for optimization. Focuses on improving RÂ^{2}, RMSE, and MAE metrics.	Industrial (Steel Industry) [43], Industrial (National) [44]
Voting Regressor Models	It employs the Voting Regressor technique for energy prediction in residential buildings using features like building type, size, and appliance usage.	Residential (Jaipur, India) [45]
Hybrid and Spatiotemporal Models	This combination enhance accuracy in residential energy consumption forecasts.	Residential (South Korea) [46]
GAN-based Ensemble Models	Ensemble model is built using GANs, stacking five models: SGD, SVC, XGBoost, LightGBM, and KNN.	Various Energy Consumption Scenarios [47]
Two-Level Ensemble Models	It uses a two-level ensemble model to predict energy demand in smart grids.	Regional Consumption [48]
Deep Learning and Random Forest Models	It combines Deep Learning and Random Forest algorithms to predict monthly energy consumption.	National Consumption (Denmark) [49]

prediction tasks and GA to optimize the hyperparameters of the model. Although multiple operational and environmental variables such as oil flow, system pressure and oil temperature were considered, the final model used only turnover as the input variable. Even though conventional variables were not fully used to calculate the energy consumption, the results of the study indicated that the hybrid SVM-GA model significantly outperformed traditional models. These results were validated by measuring different parameters such as MAPE, which was as low as 3.05 %.

Another important study that falls into the category of metaheuristic models in hybrid algorithms is the one proposed by Ref. [38] In this research, the authors use a genetic wind-driven optimization (GWDO) algorithm to optimize a hybrid model from the factorized conditional restricted Boltzmann machine (FCRBM). On the one hand, the section of the algorithm that is built by FCRBM uses autoregressive techniques to train the network, while on the other hand, the section designed with GWDO is used to optimize the model parameters, which improves the convergence rates and the prediction accuracy. Like the previously mentioned models, the authors use variables such as wind speed, dew point, temperature, humidity and historical consumption data for its construction. Finally, the researchers conclude that like several hybrid models, the proposed algorithm presents significant improvements over traditional models, with metrics such as MAPE of 1.10 %, which outperformed benchmarks such as Multiple-Input Multi-Expert Deep Ensemble–Artificial Neural Network (MI-mEDE-ANN, 2.20 %) and Adaptive Forecast Combination–Short-Term Load Forecasting (AFC-STLF, 2.10 %)

On the other hand, if the hydrocarbon industry is studied again, an algorithm built by Ref. [39] is found that calculates and optimizes energy consumption in oil pipelines, which is built by Support Vector Machine (SVM) with an Improved Particle Swarm Optimization (IPSO) algorithm. The part of the model that uses SVM works with non-linear relationships in energy consumption data, on the other hand, the part that works with IPSO is responsible for optimizing parameters to improve the prediction accuracy. Finally, the authors concluded that the model presents significant improvements over other models.

Similarly [40], proposed a PSOCA-GRNN model to predict annual energy consumption, which combines a General Regression Neural Network (GRNN) and Particle Swarm Optimization with Cluster Adaptation (PSOCA). The data used in training the model were GDP, population, and trade volumes (imports and exports). Like different hybrid models, this algorithm also had a higher performance with respect to more traditional algorithms such as PSO-GRNN and Ordinary Least Squares (OLS) linear regression. Some of the metrics found in the execution of this algorithm were: relative errors within a range of [−3 %, +3 %], with a lower MAPE: 1.29 % and a lower MSE: 2384.13. Regarding the comparison with individual models, metrics such as MAE of 6.32, R^{2} of 0.982 and a RMSE of 48.82 were found. Which can allow us to conclude that the model is highly effective with respect to simpler models to build. The algorithm was applied to forecasting energy consumption at the national level in China.

In earlier studies [41], Designed a nature-inspired metaheuristic ensemble model to predict energy consumption in residential buildings. The process, namely the Evolutionary Neural Machine Inference Model (ENMIM), integrated Least Squares Support Vector Regression (LSSVR) along with Radial Basis Function Neural Networks (RBFNN) using the Symbiotic Organism Search (SOS) algorithm in optimization of parameters. This methodology is based on the historical data of energy

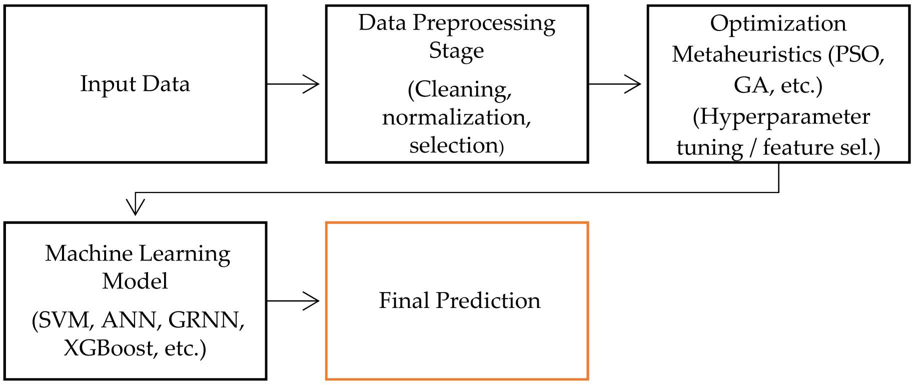


Fig. 4. General architecture of metaheuristic optimization-based hybrid models.

consumption made available from houses located in Ho Chi Minh City, Vietnam. The model achieved a RMSE of 36.31 kWh, MAE of 29.45 kWh, and MAPE of 8.90 % during testing, with R^2 of 0.93. Compared to benchmark models such as Deep Neural Networks (DNN) and Support Vector Machines (SVM), the ENMIM outperformed in accuracy and stability. With the possibility to represent complex patterns contained in the data, these symbiont models provide a useful approach to estimating energy consumption, assigning parameters to improve the estimates. The model was validated against benchmark models with the results indicating ENMIM being on par with competing approaches, which shows it as a promising tool in residential energy management and planning.

Similarly [42], presented some of the hybrid models based on stacked generalization to predict the energy consumption in the electrolytic preparation of copper foils. The hybrid PSVM-PMLP-MLR model includes Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Multiple Linear Regression (MLR) as base models. Particle swarm optimization was employed to optimize the parameters of the initial as well as the final ensemble models. So, a general procedure was executed comprising two main steps: base and meta. In the base phase, SVM and MLP were performed to study different input structures along with the optimized PSO-based combination learning to provide a lower prediction error. Then an MLR model was configured for its final prediction based on meta-learning. The experiments show that the model predicted significantly better, achieving a RMSE of 2.4397, MAE of 2.0046, MAPE of 1.5071 %, and R^2 of 0.9876. The PSVM-PMLP-MLR model delivered 10.29 % more accurate predictions than SVM and 8.28 % more accurate than MLP, which could make it a relevant tool to optimize the energy consumption parameters in industrial processes related to electrolytic copper foil production.

Hybrid models based on metaheuristic optimization have demonstrated great performance compared to traditional machine learning models. Some of these methodologies such as particle swarm optimization (PSO), genetic algorithms (GA) and jellyfish search (JS) have a great advantage in adjusting hyperparameters in addition to selecting relevant features and improving the accuracy of predictive models. Although these types of algorithms have been used in different ways in the calculation of energy consumption, the effectiveness of these algorithms depends on a balance between computational complexity and accuracy when predicting the study parameter.

Table 3 shows a summary of the main metaheuristic optimization-based models used for energy consumption prediction, categorized into three groups: PSO-based Models, Genetic Algorithm-based Models, and Symbiotic and Nature-inspired Models. Each group highlights the key characteristics of the models and their applications in various sectors, including smart buildings, industrial processes, regional energy systems, and residential energy management.

On the other hand, it was decided to compare the performance of the models analyzed in this section. As in the first part, the MAPE for each algorithm was compared. In Fig. 5, it can be seen that the most optimally performing model is the Factorized Conditional Restricted Boltzmann Machine–Genetic Wind-Driven Optimization (FCRBM-GWDO), with a MAPE of only 1.10 %, applied to the residential context. On the other hand, the least accurate model was the ENMIM (LSSVR + RBFNN + SOS), also applied to residential environments, with a MAPE of 8.90 %. This may be because the model uses a Factorized Conditional Restricted Boltzmann Machine, which is very powerful in capturing nonlinear temporal dependencies, and Genetic Wind-Driven Optimization is a very recent and powerful algorithm. This combination could be more effective than the combined use of Least Squares Support Vector Regression (LSSVR) and Radial Basis Function Neural Network (RBFNN), since these architectures do not always complement each other well. LSSVR can be sensitive to noise, and RBFNN requires precise adjustment of centers and radii. Furthermore, it is worth noting that the context of the data and its source must also be taken into account. For example, the LSSVR and RBFNN algorithms were tested in homes in Ho Chi Minh City, where energy consumption is highly heterogeneous and seasonal.

Finally, it is worth noting that the SVM model optimized with Improved Particle Swarm did not present MAPE metrics.

3.2. Ensemble models for smart grids

Ensemble models originated intending to optimize the prediction results of machine learning algorithms, generally combining methodologies such as decision trees, neural networks and support vector machines, to improve the accuracy of the results. Depending on the different combinations, ensemble techniques can be presented as bagging, boosting and stacking, which are methodologies to reduce variance, and bias and improve generalization. In the specific case of energy consumption, these techniques are useful to find intricate consumption patterns in different scenarios such as buildings, smart homes and large industrial systems.

As mentioned, some of the most representative methods described in this chapter include Bagging (Random Forest), Boosting (Gradient

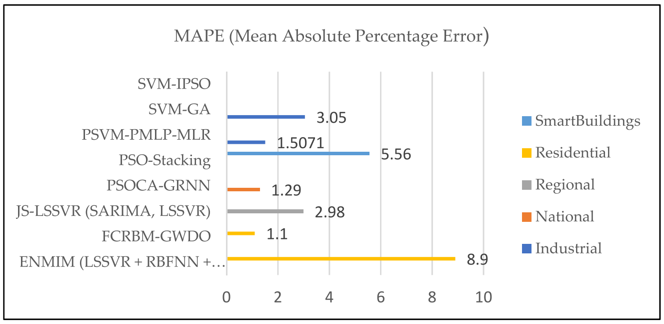


Fig. 5. Mape (mean absolute percentage error).

Boosted Trees), and Stacking (heterogeneous models). When referring to a Bagging methodology such as Radom Forest, it emphasizes a type of model that works by training different decision trees, along with parallel work using subsets of data from the original data. In turn, each tree learns independently and then combines the different predictions made. When a model is built this way, the variance is reduced, meaning the data tends to behave more homogeneously. This makes it effective for working with data with a lot of noise.

On the other hand, the Boosting methodology trains the models sequentially, with the goal of each new decision tree correcting the errors of the previous set of trees. In other words, it is a consecutive process that improves as it goes through a greater number of correction stages. To achieve this, the algorithm is assigned greater loads for erroneous predictions so that it can focus on solving these types of problems. This allows for the correction of systematic errors related to abrupt changes in demand, such as those caused by climatic phenomena or unexpected industrial variations.

A final model is called heterogeneous or "stacking," where different algorithms are trained individually—for example, support vector machines, decision trees, or neural networks—which subsequently obtain results that are used to train a metamodel, which learns to integrate all the results of the individual models to improve the prediction of the required data. This ensures that the model can adapt to the diversity of consumption patterns, in addition to being able to integrate renewable energies that are often intermittent, either due to the weather or behaviors that strictly depend on the users.

Fig. 6 illustrates the general structure of ensemble models, where multiple algorithms are combined through techniques such as stacking, voting, or averaging to enhance prediction performance.

While it is true that ensemble models significantly increase prediction accuracy by combining multiple learning algorithms, this often reduces interpretability. To counter this, we, as the progenitors of this review, used variable importance techniques, such as SHapley Additive exPlanations (SHAP) scores, gain ranking, or permutation importance, primarily in models such as XGBoost and Random Forest. These tools make it possible to identify the most relevant variables for predicting energy consumption and, consequently, improve our understanding of the model's internal behavior.

The following studies illustrate the implementation and effectiveness of ensemble models in predicting energy consumption.

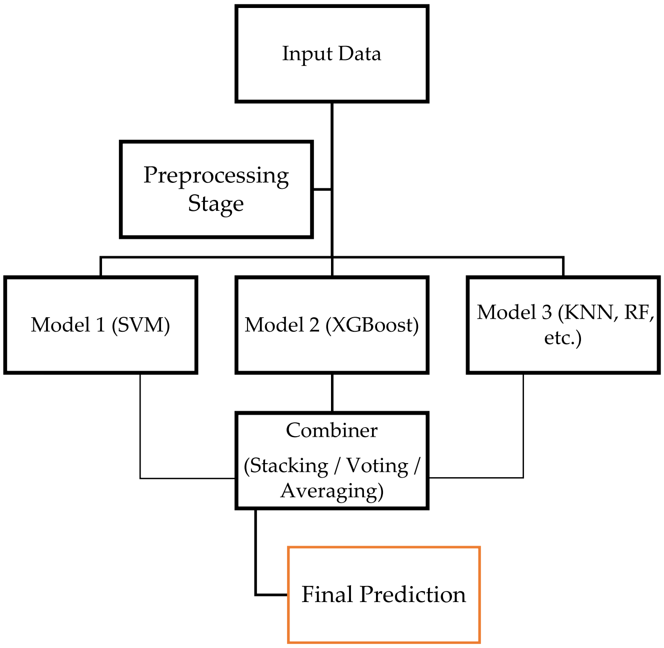


Fig. 6. General architecture of ensemble models for smart grids.

A first example of this kind of model for energy consumption calculation is the research carried out by Ref. [43] who proposed a Stack-XGBoost ensemble model for the prediction of energy consumed in the steel industrial environment, using active and reactive energy consumption data, power factors and emissions data. The algorithms implemented were base models such as Extra Trees Regressor (ETR), AdaBoost, Random Forest Regressor (RFR) and Extreme Gradient Boosting (XGBoost) as a meta-learner. As expected, the results of the algorithm yielded excellent results with very promising metrics such as an R^2 close to 0.99, as well as significant reductions in the RMSE and MAE values. The authors claim that this model can improve operational efficiency as well as reduce energy costs in this type of industry. The article does not report MAPE values in its results.

Another relevant study [44], focuses on predicting transportation energy demand in Türkiye using stacking ensemble models. The research aimed to develop a reliable forecasting system that combines 19 ML algorithms with stacking ensemble techniques to support decision-making in infrastructure policies and sustainability. The dataset included critical indicators such as GDP, population, ton-km, vehicle-km, passenger-km, and oil prices. The methodology evaluated models with and without dimensionality reduction, utilizing a 4-fold cross-validation approach. Metrics such as R^2, MAE, MSE, RMSE, RMSLE, and MAPE were used for performance evaluation. The Extra Trees Regressor-based stacking model delivered the best performance, achieving an R^2 of 0.99 when all variables were used and 0.98 with only two selected variables. The study also highlighted that dimensionality reduction and multicollinearity elimination significantly improved model performance. These results demonstrate the stacking ensemble model's robustness in handling complex datasets and its potential application in transportation energy management. The final model achieved a MAPE of 3.03 %, with an RMSE of 335.31 and an R^2 of 0.9955.

Another important study to analyze is carried out by Ref. [45], where the researchers build a model for the calculation of energy consumption in different residential buildings, based on the Voting Regressor method. To build this ensemble model, it was decided to take data from more than 3000 residences in Jaipur, India, where different variables were selected such as: building characteristics (type, size, materials and windows), use of appliances (fans, air conditioners) and occupancy and income levels. Finally, the ensemble model shows a good performance since significant metrics were known such as a MAE of 0.0191, MSE of 0.0008, RMSE of 0.0291, RMSLE of 0.0192 and R^2 of 0.9751 on the test data set. The model also achieved a MAPE of 5.0 %, confirming its strong predictive performance.

On the other hand, research developed by Ref. [46] proposed an ensemble model for energy consumption forecasts in residential buildings. The authors used spatiotemporal clustering techniques to improve accuracy; here, K-means clustering was applied to segment spatial features, in addition to combining Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to process temporal data. The data used in this research were energy consumption variables recorded over one year from smart meters installed in residential buildings in South Korea. The ensemble model performed the best and yielded a MAE of 572.693 kWh, a RMSE of 737.369 kWh, a MAPE of 4.182 % at the building level and 4.54 % at floor level and an r^2 score of 0.944. The results of the study showed that the MAE of the proposed model was improved by 18 % against that of an individual LSTM or GRU model, establishing the benchmark of effectiveness for well-timed decision--making for residential energy management and planning.

Continuing with the research review [47], proposed an improved ensemble model with generative adversarial networks (GAN), with stacking of 5 models: Stochastic Gradient Descent (SGD), Support Vector Classification (SVC), XGBoost, LightGBM and K-Nearest Neighbors (KNN). The performance of the algorithm is evaluated with the metrics MAE, RMSE and RMSE (CVRMSE). The results showed that the improved ensemble model with GAN outperformed the models that did not use this methodology, achieving reductions in MAE, RMSE and CVRMSE of 1.71 %, 1.63 % and 4.72 %, respectively.

A similar research to the previous ones presented is the one carried out by [48], who proposed a set model that presents two levels, whose objective is the calculation of energy consumption in smart networks at a regional level, in the short and long term. To train and validate the model, a UMass Smart* data set is used, which presents variables of energy use and climatic conditions. Once the model was executed, metrics that represent the usefulness and success of the algorithm could be observed, such as a precision of 95.6 %, a MAE of 0.00073 and a R^2 of 0.96.

As a final research of this section, the research presented by Ref. [49] is analyzed, where the authors propose an algorithm for the calculation and optimization of energy consumption based on the Deep Learning and Random Forest models, using data on temperature, precipitation and wind conditions, historical patterns, consumption and socioeconomic factors. The results obtained show the following metrics: MSE of 0.008 and an accuracy of 95 % for the Deep Learning model, while the Random Forest model achieved an MSE of 0.51 and an R^2 score of 98.7 %. This leads to the conclusion that the proposed model has a successful accuracy compared to other individual models.

The following table summarizes the key findings about Ensemble Models, which combine multiple predictive algorithms to enhance accuracy, robustness, and generalization.

Furthermore, in this section, it was decided to compare the coefficient of determination metrics, since most articles do not present MAPE values. Furthermore, since the research contexts are different, it is not feasible to compare values such as MAE or RMSE. Fig. 7 summarizes the value of this coefficient.

As analyzed, the Stack-XGBoost model presents an almost perfect alignment with a coefficient of 0.9981, between the predictions and the actual values in an industrial environment, however, all models have good prediction since they exceed the threshold of 0.94. On the other hand, comparing the 7 models with this parameter is limited since the importance of errors in the evaluation of a model is ignored. Fig. 7 shows a comparison of the MAPE of 4 of the 7 cited, as mentioned, the authors did not present this metric in all the articles.

Among the 4 models compared in Fig. 8, the Stacking algorithm Extra Trees (ET), Random Forest (RF), and Ridge Regression (Ridge) applied to national energy consumption presents a lower MAPE with respect to the others, this may indicate a greater prediction capacity, despite being in a complex and massive context such as the national scope. This result highlights the great capacity that this type of algorithm has to make predictions. However, in the residential context, it seems that the ensemble models are less efficient. It should be noted that the effectiveness of a model also depends on the effectiveness with which the data was taken and the context in which the research is presented.

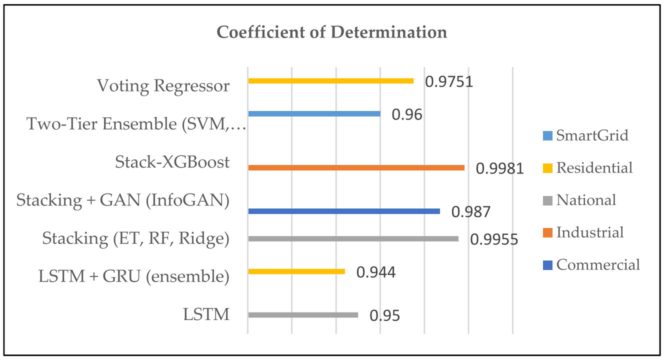


Fig. 7. Coefficient of determination.

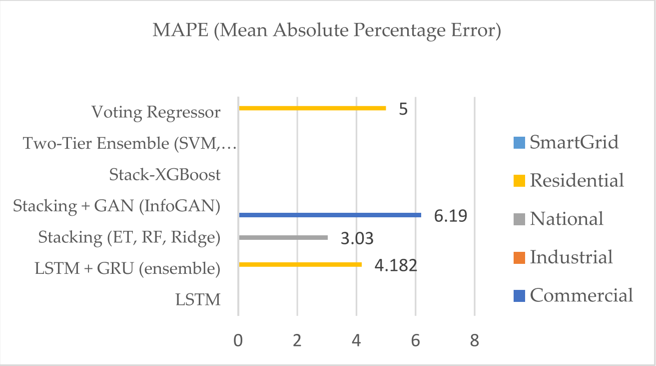


Fig. 8. Mape (Ensamble models).

3.3. *Combination of statistical, ensemble, and Hybrid Models for Smart Grids*

It has been observed that the integration of traditional statistical methods with advanced ML techniques has provided a viable approach in the area of energy consumption forecasting. Such hybrid and ensemble models incorporate conventional algorithms such as ARIMA along with SVM or LSSVR. This integration leverages the ability of statistical methods to identify linear and seasonal trends as well as to use ML to handle the challenges posed by non-linearities present in energy data. This hybrid method is highly effective in various applications including energy consumption prediction in smart grids and long-term energy demand forecasting due to its high level of flexibility for accurate use in various systems.

The use of the ARIMA technique together with LSTM lies in the fact that the first one states, wants to address separately, the linear behaviors that may present autocorrelation and seasonality, while in a second section the LSTM algorithm tries to model the complex behaviors that cannot be explained by ARIMA; thus, when the results are obtained separately, they are combined and a more optimal solution is reached in terms of energy consumption.Fig. 9 presents a general scheme of hybrid models combining statistical and machine learning components for energy forecasting.

Therefore, the following studies will be discussed by looking at the methodologies used, applications, and results of these models.

Interesting research is the one presented by Ref. [50] who propose several combined machine learning models such as XG-Boost, Genetic Algorithm, Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) to identify the most suitable set of historical climate and energy consumption variables that are best suited for building the algorithm. The performance of the various testing algorithms and the energy consumption result data from Ref. [44] revealed that the hybrid algorithm provided better improvement than the other individual algorithms. The hybrid model showed a MAPE of 3.35 %, a MAE of 10.05, a MSE of 192.32, and a RMSLE of 0.02 over a three-month testing period. The proposed model was specifically applied to short-term electricity demand forecasting for the Jeju Island power grid, demonstrating effectiveness in the complex task of predicting power demand from time series data.

Another study is analyzed by Ref. [51], in which the authors create a hybrid model combining neural networks with ARIMA models to take advantage of the different strengths of these two methodologies, since on their own they present different weaknesses depending on the nature of the data. ARIMA effectively captures linear trends in energy consumption data, while artificial neural networks (ANN) handle more complex patterns, such as non-linear relationships. The hybrid algorithm was applied to regional energy consumption forecasting using data from Hebei province in China, focusing on short-term and long-term predictions. The model demonstrated superior accuracy,

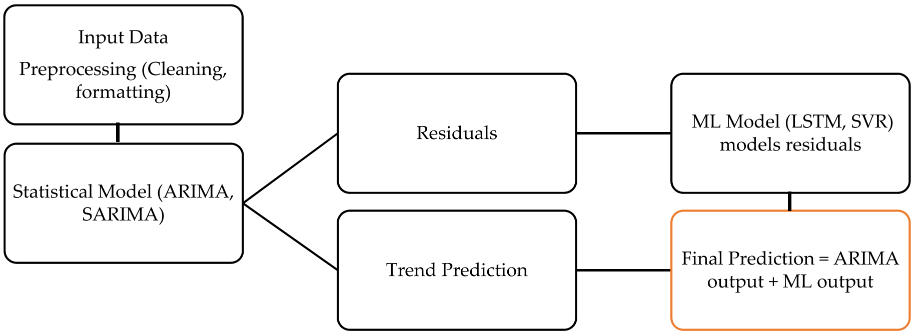


Fig. 9. General architecture of hybrid models combining statistical and machine learning techniques.

achieving a RMSE of 92.45, a MAE of 73.8, and a MAPE of 0.311 %, outperforming both standalone ARIMA and ANN models. These results highlight the hybrid model's capacity to reduce forecast errors and capture both seasonal and irregular patterns in energy consumption data.

On the other hand, a study carried out by Ref. [52] found significant differences between several models for calculating energy consumption, including statistical models (ARIMA, SARIMA), ML algorithms (Random Forest, XGBoost, Decision Trees), DL methods (LSTM, CNN, RNN), and hybrid models (LSTM-CNN, CNN-GRU). The research dealt with the context of energy load forecasting in a regional framework. It was worked with datasets that included energy consumption measurements (hourly and daily) and meteorological variables such as temperature, wind speed, and humidity collected from Spain, Germany, and the United States. The study showed that hybrid deep learning models outperformed a set of other models with variable performance in terms of time and performed scenarios. The Hybrid model LSTM-CNN performed best among all the ML and DL under test with a prediction accuracy of 92 % in the case of Germany, 96 % in the case of the USA, and 98 % in the case of Spain. In instances where augmentation techniques were used, the RMSE and MAE improved and thus validated the hybrid models with values as low as 50.15. The hybrid models, particularly those combining LSTM and CNN architectures, form the best architecture among all others for predicting region-level consumption because they capture both temporal and spatial features.

Continuing to analyze hybrid models, there is one study in particular that calculates electrical energy production using a model called H-T Transformer [53]. The authors combine two methodologies to achieve optimization and high accuracy in the final results. First, they combine the SARIMA methodology, which is very useful for working with linear patterns, with a Transformer Neural Network model, which is generally used when the data have nonlinear relationships and the goal is to reduce residual errors. In this study, the researchers compared this hybrid algorithm with other individual models, and as expected, H-T Transformer performed better than the others, with values below 25 % in the CV(RMSE) metric.

Other researchers decided to continue using the Transformer methodology to optimize the calculation of energy consumption [54]. In this experiment, it was decided to conduct an experiment with electricity consumption in a house in Sceaux (Paris, France). For this, two models were created: the first uses a combination of CNN and LSTM algorithms, and the second uses Transformer. However, before doing so, a technique called SWT is used to clean and prepare consumption signals. These models were compared with more traditional methodologies such as ARIMA, SVR, and Random Forest. Some of the results were: a RMSE of 0.1574 kW, a MAE of 0.1106 kW, and a MAPE of 11.72 %. This leads to the conclusion that this hybrid model is more accurate in the weekly resolution than the traditional algorithms with which it was compared. To better illustrate the training performance of these models, Fig. 9 shows the convergence curves of both the CNN-LSTM-SWT and Transformer-SWT architectures across 100 epochs.

Fig. 10 shows how the prediction error decreases over training epochs, reflecting effective convergence of the hybrid models.

Another relevant study to analyze is the one developed by Ref. [55] here the researchers propose a hybrid model that is based on federated learning which uses time series features in addition to a clustering algorithm, among which are K-means++ and Affinity Propagation, this is done in order to calculate the energy consumption in different homes. If compared with individual models, federated learning achieves notably lower error rates and in addition to this it can maintain high data privacy and reduce the communication load. As for clustering techniques such as Affinity Propagation, they can improve prediction and performance since it can group consumers into different profiles. Among the most notable values of the results are lower RMSE metrics as low as 0.5473 and an average MAE of around 0.3517 in the different clusters.

On the other hand, researchers from [56], propose a hybrid model which also tries to calculate and optimize energy consumption. This is done by integrating several machine learning algorithms, among the main ones are: Auto-regressive (AR), Multilayer Perceptron (MLP), Extreme Learning Machines (ELM), Radial Basis Function (RBF) and Echo State Networks (ESN), with ELM serving as the main combination model. In order to train and validate this model, different measurements of energy consumption and the historical behavior of the data were made. Specifically, this research focused on analyzing data from different homes. The final results yielded successful metrics such as concordance index of 94.1 %, with MAE values around 0.0257, RMSE of 0.0434 and a MAPE of 13.54 %.

In a different context [57] proposed a system for the prediction of short-term energy consumption for smart cities, implementing hybrid and ensemble models such as LSTM, XGBoost, and LSTM-Transformer. Among these algorithms, the authors conclude that the hybrid

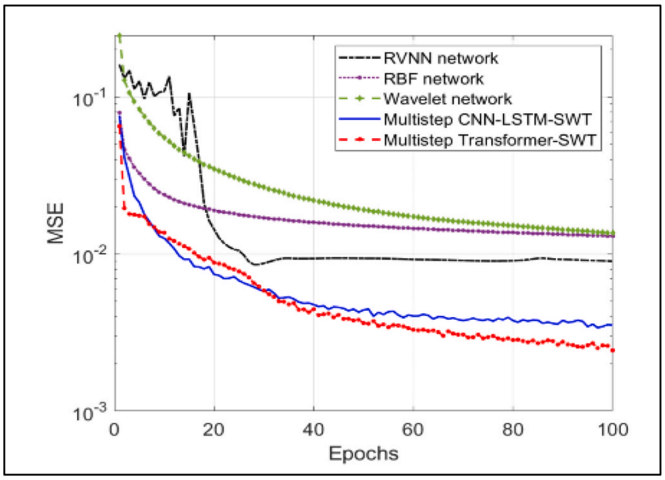


Fig. 10. Convergence curves of CNN-LSTM-SWT and Transformer-SWT models during training over 100 epochs. Reproduced from Ref. [54].

LSTM-Transformer model achieved the best performance for industrial energy consumption in Berlin, with an RMSE of 0.087, MAE of 0.053 MW and a MAPE of 0.57 %. The dataset referred to as JERI-CHO-E usage is characterized by industrial sectors, including some of the following important variables: time (hour, day, month), lighting energy factors, and some ICT data. This research is particularly applied to industrial, commercial, and residential environments and therefore is relevant for urban environments in terms of smart grid context.

Continuing with the model exploration [58] created a hybrid algorithm that uses tools such as XGBoost, CatBoost, and Random Forest to improve prediction accuracy by first imputing missing values in the dataset before applying the predictive models. The methodology focuses on imputing missing or incomplete values to improve the reliability of the results. The hybrid model achieved significant improvements in prediction accuracy, with an R^2 of 0.9212, outperforming individual models such as XGBoost (0.8195) and CatBoost (0.8294) and reaching a MAPE as low as 0.81 %.

Another interesting analysis model is proposed by Ref. [59] who developed a hybrid algorithm for China's energy consumption. The model integrates the Group Method for Data Handling (GMDH) with AdaBoost ensemble techniques and nonlinear prediction models including Backpropagation (BP) neural networks, Support Vector Regression (SVR), Genetic Programming (GP) and Radial Basis Function (RBF) networks. This multi-model effectively outperformed 7 other more basic models in terms of efficiency and accuracy. Some of its key metrics were as follows: MAPE as low as 1.20 % and RMSE of 0.4672. The relative error remained mostly between −3 % and +3 % as shown in the error plots. In addition, the metrics of MAPE and MSE showed improvements compared to traditional models.

A different study by Ref. [60] analyzed energy consumption in China, using a combined model based on ARIMA(2,1,1) and GM(1,1) methodologies. Some of the observations are that the ARIMA(2,1,1) model performed well for long-term trends, while GM(1,1) was able to perform better with recent data, despite this, the model performed well, achieving a mean absolute percentage error (MAPE) of 2.30 %, compared to 4.62 % for ARIMA and 3.75 % for GM(1,1). The article points out that thanks to this model it was possible to show that energy consumption in China is growing at a slower rate, which is important to implement different strategies regarding this energy consumption.

Another hybrid model studied is the one developed by Ref. [61] who evaluated different machine-learning techniques to calculate energy consumption in residential buildings using real data from an experimental smart grid. The research performs simulations with simple, hybrid, and ensemble models. In this global comparison of models, it is determined that models with a hybrid approach are more accurate than individual models or even ensemble algorithms, achieving a MAPE of 15.657 %, with MAE values around 0.028 and RMSE values as low as 0.164. This finding is important since a hybrid model can be directly selected if what is desired is efficiency and accuracy when calculating energy consumption. It should be noted that the selection of these models is also based on the intrinsic nature of the data; therefore, preliminary statistical studies must be carried out and the context of the research must be understood to select the appropriate model.

The following table summarizes the key findings about Statistical, Ensemble, and Hybrid Models for Smart Grids, highlighting their integration to capture both linear and nonlinear energy consumption patterns.

These Statistical, Ensemble, and Hybrid Models are particularly effective in handling energy consumption forecasting in structured datasets but face challenges in adapting to highly volatile or incomplete data scenarios.

As in the previous sections, the MAPE was compared for this type of models. Fig. 11 shows that the best model is the LSTM + Transformer, applied in industrial environments, with a MAPE of 0.57 %. This could happen because this type of algorithm can work with time and nonlinear patterns that present high complexity. In addition, as mentioned above, data taken in industrial environments can have more stable patterns since these processes are often automated. On the other hand, the hybrid Ensemble Learning model with Extreme Learning Machine (ELM) was the one that had the worst performance with 13.54 % applied to the smart buildings sector, this could happen due to the complexity of the factors influencing energy consumption in these environments, which can include variables such as occupancy patterns and energy efficiency systems.

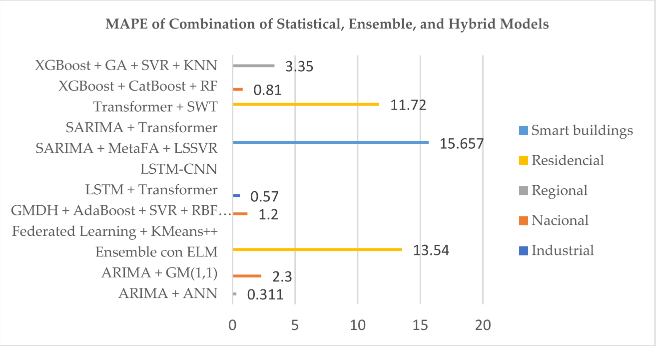


Figura 11. MAPE of combination of statistical, ensemble, and hybrid models.

3.4. *Hybrid and ensemble ML models for consumption forecasting: applications to microgrids*

A microgrid is a decentralized electrical network that integrates different energy sources and has local storage and consumption. This microgrid can be connected to a larger electrical network or it can also operate independently. The use of this type of network allows resources to be managed more efficiently, whether renewable or non-renewable since they can improve the resilience of the electrical supply and can also reduce losses in the transmission process.

This section will examine different investigations in which the authors use hybrid and ensemble models to predict energy consumption in this type of microgrids, where they combine the efficiency of multiple algorithms to improve accuracy in complex situations where there are different types of significant variations caused by the integration of renewable sources, such as solar and wind.

The first study analyzed in this research is the one published by Ref. [62], which focuses on the prediction of energy consumption in this type of system. The main objective of the work is to propose a hybrid model which is based on the use of self-adaptive fuzzy logic and is also optimized through evolutionary strategies and the GRASP procedure. The data used by the researchers are classified for residential and commercial microgrids. Once the algorithm was developed, it was tested on real microgrids to obtain different results, which showed that the MAPE was less than 10 % in most cases. In particular, MAPE values of 5.039 % and 7.029 % were obtained for microgrid D, which shows that the model improves the accuracy of the calculations despite the high variability of the data. In terms of RMSE, the results were also competitive, with values such as 13.162 MW and 16.456 MW for different validation weeks.

Another relevant study that was chosen for the analysis is the one published [63] where researchers analyze the use of machine learning or artificial intelligence (AI) models to improve the integration of microgrids with renewable sources such as solar, wind and water sources such as ocean waves. To meet the objectives of the research work, it was decided to perform simulations in the MATLAB software with the Simulink application, where an artificial neural network algorithm (ANN) was specifically used, which was combined with particle swarm optimization (PSO), seeking that the integration of these two models would improve the prediction of the simulations in addition to the optimization of the processes. The results of the simulations managed to obtain a Normalized Mean Square Error (NMSE) of 1.10 % and a MAPE of 5.04 % in 3367.50 s, evidencing an accuracy of 98.9 % in the predictive models. The study concludes that AI significantly improves microgrid management by optimizing demand prediction.

Another relevant study is the one published by Ref. [64], where researchers focus their work on predicting energy demand in microgrids, specifically with highly variable loads. To do this, the authors evaluated 15 prediction methods that include statistical models, machine learning (ML) techniques, and ensemble methods, which were compared to each other. To apply the simulations, energy consumption data from two circuits of a wastewater treatment plant were used, which were characterized by high changes in demand over time. As expected, in the analysis of results, it was evident that the best method was Support Vector Regression (SVR), which achieved precision values higher than 92 % (R = 0.9269) a reduced RMSE, and a MAPE of 61.78 %. This high MAPE value is mainly due to the highly variable and unpredictable nature of the power demand data analyzed.

In another context, the research carried out by Ref. [65] focuses on the study of urban microgrids based on renewable energies, specifically on the prediction of energy consumption. For this, the authors developed a hybrid model that combines the LSTM and SARIMAX algorithms whose objective is to improve the accuracy in the prediction of the electric load and wind power generation. To carry out the simulations, the researchers used historical data of variables such as electric consumption and wind speed in Montreal, Canada. In the final results, it was observed that the MAPE for load forecasting was reduced to 7.36 %, and the RMSE to 44.45, showing a significant improvement compared to the individual models. Similarly, for wind forecasting, the RMSE was reduced by 10.5 %–16.6 %, compared to the use of isolated prediction models that did not apply algorithmic combination techniques.

Another particular study is the one developed by Ref. [66], which designs a hybrid model for forecasting energy consumption in low-carbon microgrids. The methodology proposes combining highly complex data decomposition techniques, such as the decomposition of empirical modes in sawtooth ensembles (Ensemble Empirical Mode Decomposition - EEMD) and time series prediction, such as Convolutional Long and Short-Term Memory (CLSTM) neural networks, which were optimized by the Grey Wolf Optimizer (GWO) algorithm. This research used two sets of real data on household energy consumption, the first one made up of 200 homes with records every half hour and another one of 300 homes where the record is made every hour. Once the simulations were carried out, the results obtained showed a reduction in RMSE by 80 % compared to traditional models and an improvement in R^2 of 3 %. The best model achieved a MAPE of 4.16 %, which reflects excellent short-term forecasting accuracy.This accuracy allows microgrids to anticipate energy consumption with much greater precision. All of this helps to integrate renewable energy, as well as contributing to the emergence of demand-side management methodologies, such as dynamic pricing and demand response programs.

Finally, significant work is the one carried out by Ref. [67], who built a hybrid model for short-term energy consumption prediction in a solar microgrid using data from advanced metering infrastructures (AMI). The model makes use of convolutional neural networks (CNN) and bi-directional gated recurrent units (Bi-GRU), which are optimized by hyperparameter tuning. To achieve this, the data considered are electrical consumption, solar energy production, battery storage and a set of meteorological variables collected every 15 min in Taiwan for 14 months. The model obtains improvements of up to 71.88 % and 49.54 % in the values of MSE and MAE, respectively, compared to other models such as RNN, LSTM and GRU, which allows optimizing the management of the microgrid.

As can be seen, both hybrid and ensemble models are highly effective in the context of microgrids, especially those that combine algorithms containing architectures such as LSTM and CNN, which make the process more assertive and better optimized, even facilitating the integration of renewable energy sources. Specifically in this microgrid system scenario, where stability and energy efficiency depend on rapid and accurate responses, these models can improve these processes thanks to their ability to make real-time energy consumption predictions.

Fig. 12 shows how the Ensemble Empirical Mode Decomposition–Convolutional Long Short-Term Memory–Grey Wolf Optimizer (EEMD-CLSTM-GWO) model, with a MAPE of 4.16 %, applied to low-carbon residential microgrids, performed best compared to the others. This set of techniques appears to have been successful in making the model perform optimally. On the other hand, the Support Vector Regression (SVR) algorithm applied to industrial microgrids achieved a MAPE of 61.78 %. This could be due to the extreme unpredictability and volatility of the demand data recorded every 10 s.

3.5. Validation methods in the reviewed studies

When performing these types of simulations with artificial intelligence, it is important to ensure that the results are valid, that the outputs have reliable values, and that these models can be used in different scenarios in the future. Therefore, proper validation is important.

To achieve this, studies generally use three types of methodologies: rain-test split, walk-forward, and out-of-sample. When the "rain-test split" technique is used, it refers to using part of the data for training and another part for validating the results. This validation is generally done comparatively, ensuring that the calculated data are similar to the real data. On the other hand, the "walk-forward" technique uses progressive training windows to make predictions based on the data. Finally, the out-of-sample methodology validates the procedure by taking into account a different time sequence between the training and comparison data; that is, it consists of training the model in one data period and validating it in a subsequent time period not used during training. It should be noted that these three categories were mostly used by the different authors cited in this study.

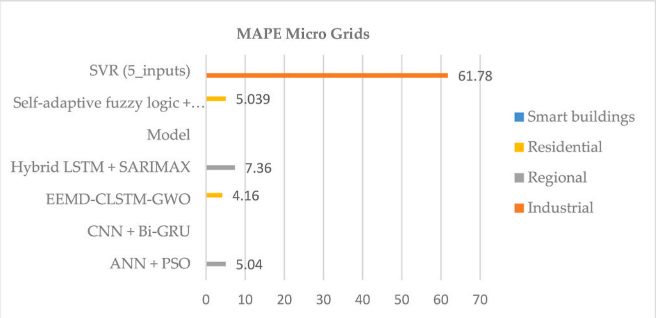


Fig. 12. Mape Micro grids.

4. Discussion

Hybrid models have shown high efficiency compared to traditional or individual models, proving that they can solve problems where the variables have even non-linear behaviors. The precision that these types of models manage has been developed in different scenarios, such as smart buildings and industrial complexes. The review of the different studies shows that predictive models based on hybrid and ensemble techniques offer better performance in terms of accuracy and absolute error in the field of smart grid applications. In this sense, the results offered by ensemble models, such as Random Forest and Gradient Boosted Trees, and models based on neural networks, such as LSTM and Bi-GRU, reveal accuracies greater than 90 % and reductions in RMSE of up to 80 % compared to traditional statistical methods, respectively. As for absolute errors, the MAPE relative errors obtained in most cases are less than 10 %, which shows that they are capable of handling the variability of the energy load and changes in demand efficiently.

Ensemble models have shown significant improvements in reducing prediction errors and increasing robustness by integrating multiple base algorithms. This happens because by combining different methods they can capture complex patterns in the energy system in different scenarios. However, one of the limitations is the computational cost and the need to obtain abundant and high-quality training data.

While complex preprocessing techniques such as temporal analysis and data augmentation using GANs were a game-changer, hybrid and ensemble models have greatly benefited from these techniques. Furthermore, methods such as feature selection, recursive feature elimination (RFE) or clustering have also been shown to additionally contribute to improving accuracy and noise in the dataset. Data quality and proper identification of independent variables remain important aspects for the success of these approaches.

Another limitation observed in this research is the difficulty in generalizing the results of each model, since many of these studies are applied to specific contexts with very small sample sizes, such as homes, industrial complexes, and smart buildings (e.g., the UMass Smart* dataset, smart homes in the United Kingdom, industrial plants in India). Because of this, it is difficult to objectively extrapolate the results to more general energy consumption contexts orto make valid comparisons between the different case studies. Furthermore, the data used to train the models sometimes face privacy issues, since in many sectors these data are not public, this problem complicates the reproducibility and comparability of the results. Therefore, to improve future analyses, we propose promoting the use of open-access reference data, in addition to developing standardized protocols for the collection, analysis, and development of each model. This could ensure better comparability in calculating energy consumption.Both hybrid and ensemble models are suited to different contexts in consumption calculation, from residential networks, and smart buildings to industrial complexes. The challenge for researchers is data collection, especially at national or regional scales where data comes from multiple sources and must be in a specific structure with certain statistical characteristics so that models can be trained. Researchers need to take into account the availability of data in order to propose different solutions with these machine-learning models.

When discussing the selection criteria for the correct model to use, it is important to mention the nature of the data itself, for example, when energy consumption is highly influenced by non-linear factors and that have heterogeneous properties such as meteorological, social and economic variables, hybrid models are the most effective to capture this type of interactions that are often complex and do not necessarily have linear behaviors, this happens because each model that is part of the hybrid algorithm has different strengths that can be combined to solve this type of problem. On the other hand, ensemble models are better suited to work with data that are robust and tend toward homogeneity.

It can be observed that the vast majority of complex models, whether hybrid or ensemble, improve measurement accuracy compared to traditional machine learning models. However, it is important to highlight that there may be a risk of overfitting, especially in scenarios with

Table 5
Statistical, ensemble, and hybrid models for smart grids.

Studied Models	Characteristics	Applications
Machine Learning & Hybrid Models	The combination of these algorithms is done with the aim of improving the accuracy of the calculations.	Smart Grids for regional, residential, and smart buildings [50,51,61]
Deep Learning Models	It Transformer-based models to capture temporal and spatial patterns	Smart Grids for regional and industrial consumption [52,57]
Federated and Clustering-Based Models	It enhance data privacy and improve prediction performance.	Smart Grids for residential consumption [54,55]
Time Series & Combined Statistical Models	It capture long-term trends and recent data changes.	Smart Grids for regional energy consumption [53,60]
Hybrid and Ensemble Models	It focuses on maximum data optimization and error reduction.	Smart Grids for residential and industrial environments [56,58,59]

Table 6
Hybrid and Ensemble ML models for consumption forecasting: applications to microgrids.

Studied Models	Applications (Smart Grid Context)
Hybrid Model with Fuzzy Logic	Residential and commercial microgrids [62].
ANN + PSO	Smart buildings [63].
Ensemble Methods	Industrial microgrids [64].
LSTM + SARIMAX	Urban microgrids and renewable energy [65].
CLSTM + EEMD + GWO	Residential microgrids [66].
CNN + Bi-GRU	Smart solar microgrids [67].

Table 7
Validation and regularization techniques employed across all reviewed studies.

Validation/Regularization Technique	References (Article No.)
Train-test split/Hold-out	[17–30,35–40,42–53,55–62,64–67]
k-fold cross-validation/Bootstrap	[31,32,35,37,41–46,67]
Dropout	[31,66]
Hyperparameter optimization (metaheuristic/grid/manual)	[23,31,32,35–42,44–46,49,50,56,61,62,67]
Signal decomposition (EMD, SSA, Wavelet, etc.)	[26–29,32,66]
Data augmentation (GAN, InfoGAN, cGAN, etc.)	[47]
No robust validation/regularization reported	[33,34]

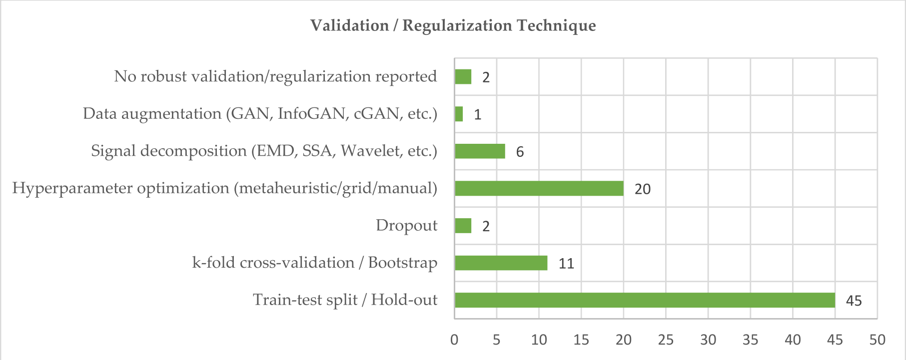


Fig. 13. Frequency of validation and regularization techniques reported in the reviewed studies.

small and noisy data; this can be assessed with some validation techniques described in the literature. In this review, some authors present conventional methodologies for performing this evaluation, such as the train-test split, walk-forward validation, or out-of-sample validation. Some authors even use more rigorous techniques such as k-fold cross-validation, in addition to metaheuristic optimization and feature selection techniques (e.g., PSO, GA, GWO), which are widely used for hyperparameter tuning but do not replace comprehensive validation for overfitting. Considering the above, it is pertinent that future research structures more advanced validation methodologies to mitigate the risk of overfitting in the different models. It is also recommended that authors report, in a clearer structure, the methodology used to solve this problem that arises in robust algorithms such as hybrid and ensemble algorithms. A summary of the main validation and regularization techniques used in the reviewed studies is presented in Table 7 and Fig. 13 below.

The most commonly used methodology is the simple train-test split, while techniques such as dropout and data augmentation are the least frequently applied across all reviewed studies.

Another important criterion to analyze when selecting an appropriate model for calculating energy consumption is the stratification of the time series since hybrid models such as LSTM or Transformer are more effective when the time interval is too small, such as minutes, seconds, or milliseconds, while combined statistical algorithms such as ARIMA-XGBoost are better suited to monthly or annual time series. This is important to mention since data often present different configurations in terms of time, and the researcher must adapt his calculations to this data configuration.

In more robust systems such as national or multi-energy systems, a good model must be selected that adapts to this type of configuration, for example, ensemble techniques combined with evolutionary optimization such as PSO-Stacking adapt well to this type of data, however, it must be taken into account that the implementation of these models in real-time has great computational limitations, which is why they present new challenges for researchers, where new hybrid models are proposed that combine computational efficiency with the precision of the calculation of energy consumption.

In spite of hybrid and ensemble models like CNN-BiGRU, transformers, stacking, and federated learning achieving previously unlikely predictive accuracies, a full review of the literature shows that scaling and real-time applicability remain formidable challenges to the actualization of these solutions in the energy sector. None of the studies evaluated implements or validates their models in an edge computing environment or on resource-constrained embedded hardware(e.g., IoT smart meters or residential controllers). All tests and evaluations are performed in laboratories, personal computers, or high-performance servers, focusing solely on predictive accuracy metrics such as MAPE, MAE, or RMSE, while bypassing any investigation of latency, inference times, and computational energy use, this presents a significant gap between academic development and real-world implementation, especially when much of this research—because of its complexity and computational demands—may prove unfeasible to implement on low-power devices, which are the standard in edge computing and IoT infrastructure. Although federated learning shows promise in addressing privacy concerns and data distribution, the majority of papers present only simulations without experimental testing on hardware.

In future research, this type of algorithm can be implemented taking into account other types of independent variables, for example, optimizing the calculation of energy consumption in specific countries, taking into account renewable and non-renewable energy production, in addition to variables such as imports, socioeconomic behavior, among others.

On the other hand, in future research, combinations of hybrid models with ensemble models could be explored to further improve the accuracy of the calculation of energy consumption; however, it is a challenge since these types of combined models require a high computational cost. However, it could be hypothesized that the prediction would be above any hybrid or ensemble model on its own.

4.1. Ethical, privacy, and implementation concerns in AI-based energy forecasting

In these types of machine learning models, data must be collected from both energy consumers and different operations, which are susceptible to privacy violations. It is important that these investigations be conducted in compliance with the General Data Protection Regulation (GDPR) in Europe, which requires standards for consent, data minimization, and data access rights to address cybersecurity threats. It is essential to implement robust encryption, authentication protocols, and continuous monitoring to ensure data privacy and security.

Furthermore, these types of models can amplify biases present in the data, which can lead to unfair or discriminatory results, affecting vulnerable populations. Therefore, it is important to identify and mitigate these problems from the outset. Furthermore, some algorithms are trained with data that is not open to the public, or that is partially private, and this does not allow for adequate transparency in the results obtained. Therefore, it is important for developers to implement equity metrics, conduct audits, and work with public data to ensure that the results are reliable and not skewed toward particular interests or biases.

Specifically in the energy field, bias problems generated in the construction of these machine learning models can include, for example, incorrectly assessing energy needs or credit risks, which can lend themselves to the manipulation of energy rates. Furthermore, there is also the risk of exposure to cyberattacks that disrupt the proper functioning of the energy system, in addition to influencing government decision-making regarding energy consumption and production.

As mentioned, it is important that, when manipulating data and building these types of models for the energy sector, recognized ethical frameworks be established to guide the responsible development and implementation of AI technologies, initiatives such as the IEEE Ethically Aligned Design and the European Union's Artificial Intelligence Act can be taken into account when building these types of models in the future, to ensure complete transparency and security in both the training of the algorithms and the results obtained.

5. Conclusions

The choice between a hybrid model or an ensemble model for calculating energy consumption depends on the objectives set out in the research since hybrid algorithms are more effective when capturing linear and non-linear patterns, as well as seasonal trends and fluctuations that tend to be unpredictable. Ensemble models are more commonly used to improve accuracy since the combination of multiple algorithms makes the data fit towards a solution as accurately as possible; However, these ensemble models have a problem that has to do with the computational complexity they require. Knowing this, an analysis of objectives, the nature of the data and computational resources must be raised in order to select the appropriate model for calculating energy consumption.

Data quality and the appropriate selection of the multiple machine learning model is a fundamental action for energy consumption predictions to be highly effective in all simulations, for example, techniques such as temporal decomposition, data augmentation with GAN and redundancy elimination using RFE are better in the accuracy of predictions compared to simpler or individual models, however, these strategies must be adapted to the characteristics of the data and the context of the research, since not all models adapt to all possible existing situations.

Both ensemble models and hybrid models have been adapted to different problematic situations in the field of energy consumption, however, at a national level, further advances are required regarding scalability and computational efficiency, since these simulations in real-time present several computing challenges, however, the use of techniques such as attention mechanisms, evolutionary optimization and deep learning could serve to overcome these limitations without compromising the efficiency of energy consumption calculation.

The studies analyzed show that energy consumption is highly influenced by different variables of a non-linear nature such as temperatures, climatic states, socioeconomic dynamics and energy production, therefore, it is strictly necessary that the prediction of this parameter be modeled with advanced algorithms that can work with this type of variables, since both hybrid models and ensemble models can address the different challenges presented by these specific data configurations.

The results obtained from the studies analyzed show that the use of advanced algorithms such as neural networks, LSTM, GRASP, Random Forest, Bi-GRU, etc. allows for the reduction of error metrics such as MAPE, RMSE, NMSE, achieving accuracies of over 90 % in many cases. Thanks to these metrics, it can be seen that this type of model is very important when talking about stability in a Microgrid, as well as the transition towards systems with low levels of carbon emissions.

Temporal variables have a significant effect on the calculation of energy consumption in machine learning models, such as hours of the day, days of the week or seasons, since algorithms that integrate these variables in their construction tend to have greater precision, since they take into account peaks and fluctuations that are only seen when analyzing this type of variability over time. Among the models that are best optimized with this type of temporal parameters are hybrid algorithms with temporal decomposition or those based on attention mechanisms.

The analysis of the reviewed studies shows that applications of predictive models in smart grids and microgrids are mostly represented by the residential sector, with around 43 % of the studies, followed by the industrial sector, which covers approximately 26 %. Meanwhile, applications in smart buildings represent around 13 %, while the regional energy consumption category corresponds to 19 % of the studies analyzed.

Although hybrid and ensemble models have shown very good results in research, several practical problems arise when trying to use them in the real world. For example, making real-time forecasts, which is very important for smart grids, requires models to update very quickly (with low delay) and learn from new data almost immediately, and this is not yet fully achieved with current models. Another problem is that the actual data used in the real world are not as perfect as those in studies: there are missing data, measurement errors in sensors, and data that come from very different sources (heterogeneous). All of this can significantly harm the accuracy of models when they move from laboratories or simulations to real-world use in companies or power grids.

When working with explanatory variables that are quite fluctuating and with inherent intermittencies, such as solar, wind energy production and climate, studies show that hybrid models that combine deep neural networks with evolutionary optimization techniques are superior in accuracy and optimization compared to simpler or individual models, since they can handle these uncertainties so as not to neglect the predictive capacity and stability of the algorithm.

Prior literature does not fully assess the feasibility of implementing these advanced models on hardware-constrained devices. For model frameworks to appeal in the marketplace, it is critical that future work focus on computational efficiency, reducing model complexity, and experimentation on actual physical platforms to confirm real scalability and applicability within smart energy systems that function in real time and under edge computing and IoT conditions.It is important to develop new methodologies that provide solutions to computing time problems, for example, the creation of new large-scale data acquisition methods, integrating new technologies such as sensors and Internet of Things devices. In addition, new work must be carried out proposing algorithms that work in real time, which as data is supplied, the model can perform the relevant calculations, in order to improve the applicability of hybrid and assembly models, running for example in dynamic processes such as smart grids and similar processes.

As consulted and described in this review, a greater number of hybrid models have been used than assembly models. This can happen for two things, first, it may be due to the low computational cost of a hybrid model compared to an ensemble model, and second, hybrid algorithms can handle data with different problems in their nature, such as non-linear relationships, behaviors, abnormalities and missing data.

Author contributions

Conceptualization, V.H.R.C and L.M.; methodology, L.M.; validation, L.M. and M.A.-O; formal analysis, V.H.R., L.M. and M.A.-O.; investigation, V.H.R.C.; writing—original draft preparation, V.H.R.C. and L.M.; writing—review and editing, M.A.-O. and P.S.; visualization, L.M., M.A.-O. and P.S.; supervision L.M., M.A.-O. and P.S. All authors have read and agreed to the published version of this manuscript.

Ackoledgments

Funding for open access charge: CRUE-Universitat Politècnica de València.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: nothing If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Case of application

Scopus and Web of Science have been chosen as reference databases for this research, the reason relies on the relevance of these two databases. Scopus covers more than 24000 peer-reviewed journals, conferences and books from worldwide similarly Web of Science covers more than 25000 journals from 80 countries. In addition to the wide coverage of high-quality research in the fields of engineering and technologies, these two reliable platforms both provide data on citations, fields of the research article and a wide variety of search metrics to evaluate the impact of the research articles. Taking into account that, other databases have been excluded from the analysis.

To ensure the relevance and quality of the articles retrieved from these databases, a pre-screening process was implemented. This process involved establishing specific criteria to filter articles, focusing on aspects such as publication years, document type, accessibility, and the inclusion of studies reporting quantitative performance metrics. The details of these criteria are summarized in Table 8.

Table 8
Screening process and selection criteria

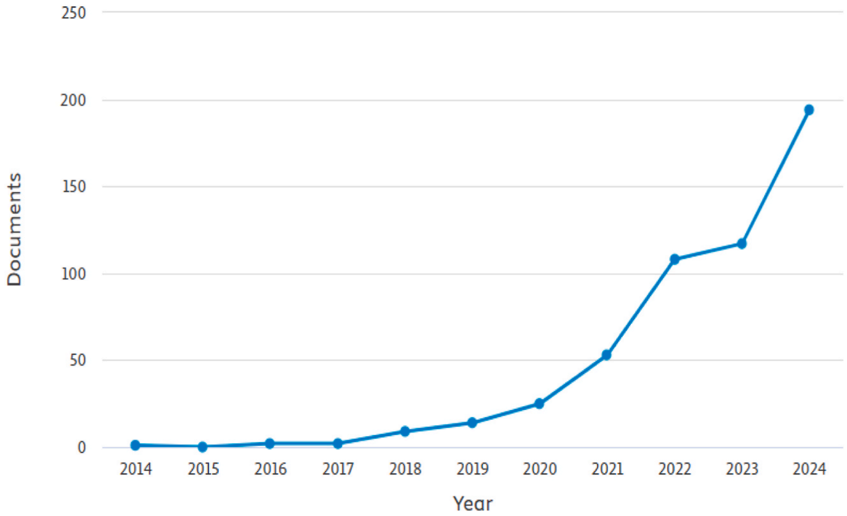
Screening process	Criteria	Details
Pre-screening Phase	**Publication Years**	2012–2024
	Max date of Publication	December 2024
	Topic keywords	Hybrid ML models Ensemble ML models Multi-level ML ensemble models Energy consumption forecasting Energy demand prediction Energy load forecasting
	Metrics	Articles reporting quantitative performance metrics (e.g., RMSE, MAE).
	Document Type	Included: Research article; Review article. Excluded: Conference proceedings, theses, and grey literature.
	Accessibility	Open Access

As can be seen in Table 8, different keywords were selected that are representative of the topic investigated in this review, which were used in the Scopus and Web of Science search engines. These words were classified as follows:

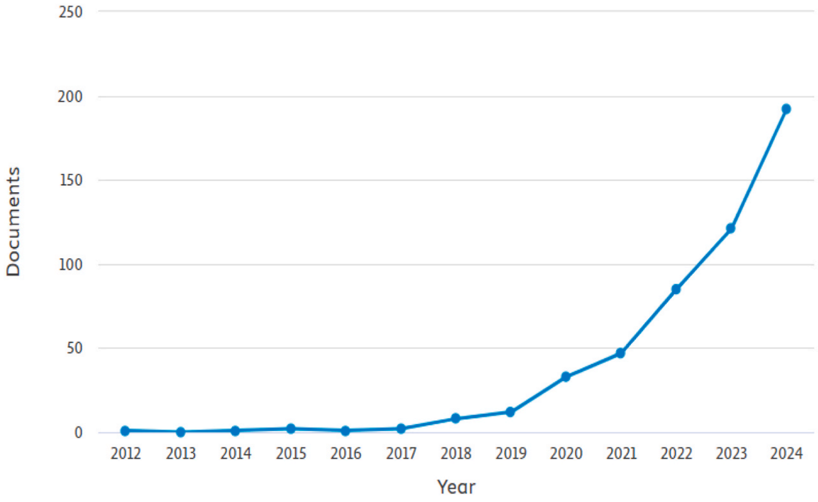
Hybrid and ensemble machine learning models: " Hybrid Machine Learning Models", " Ensemble Machine Learning Models", " Multi-level Machine Learning Ensemble Models".

Energy prediction: "Energy consumption forecast", "Energy demand prediction", "Energy load forecast"

To visualize the relevance and growth of these keywords over time, the frequency of publications associated with each term was analyzed individually. The results are presented in grouped figures, where the first set of graphs illustrates the trends for keywords related to hybrid and ensemble machine learning models, and the second set focuses on energy prediction terms. This analysis, conducted using the Scopus database, provides insights into the interest and applicability of these topics within the scientific community over the period 2012–2024.



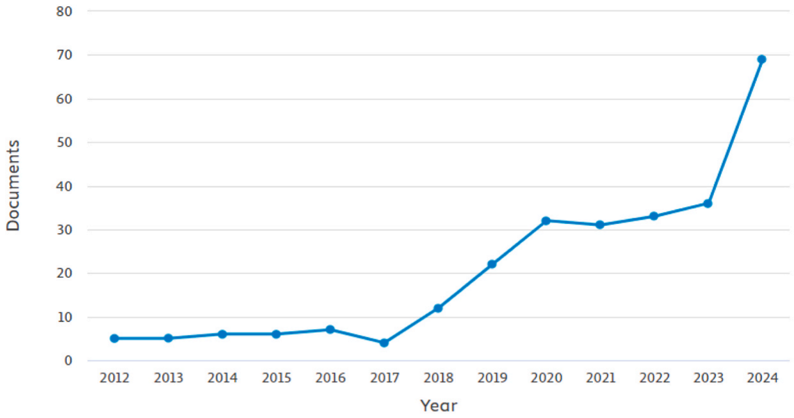
(a) Keyword search: Hybrid Machine Learning Models (525 documents)



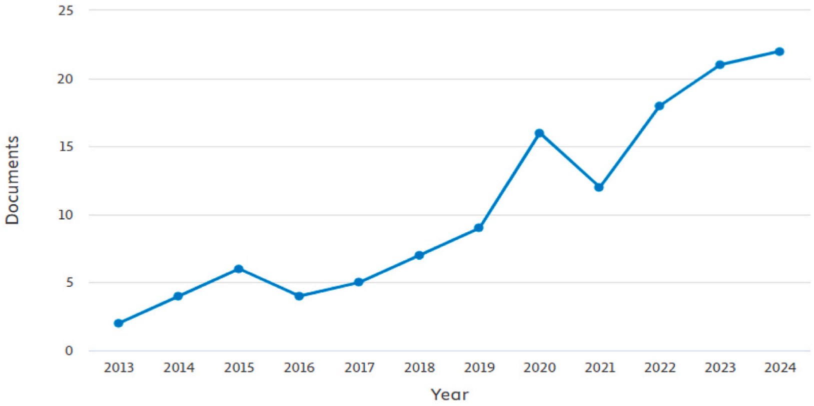
(b) Keyword search: Ensemble Machine Learning Models (505 documents)

Fig. 14. Frequency of publications for keywords related to hybrid and ensemble machine learning models (2012–2024).

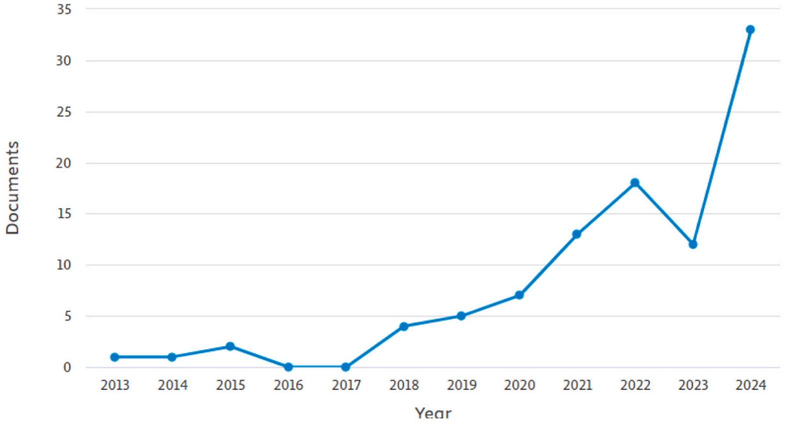
Fig. 14 illustrates the annual frequency of publications for each keyword related to hybrid and ensemble machine-learning models. Panel A highlights the increasing interest in hybrid models, while Panel B shows consistent growth in ensemble models. However, for "Multi-level Machine Learning Ensemble Models" (Panel C), no results were obtained in Scopus, indicating that this specific term has not been widely adopted or explored in the scientific literature within the analyzed timeframe.



(a) Keyword search: Energy consumption forecasting (268 documents)



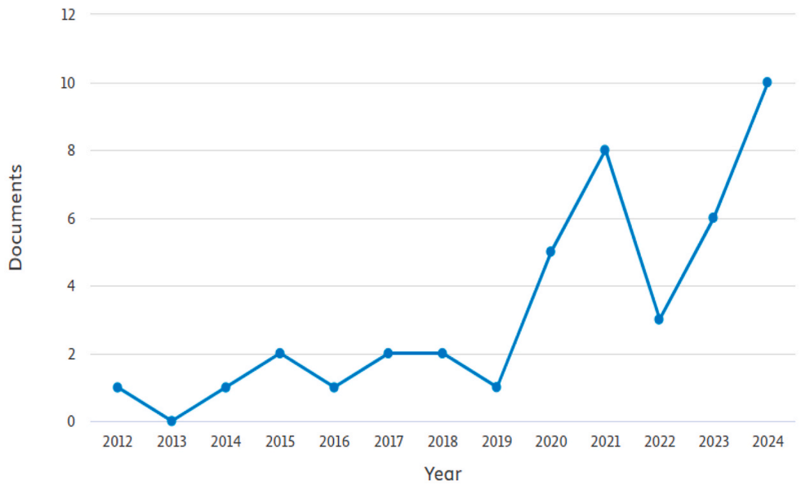
(b) Keyword search: Energy demand prediction (126 documents)



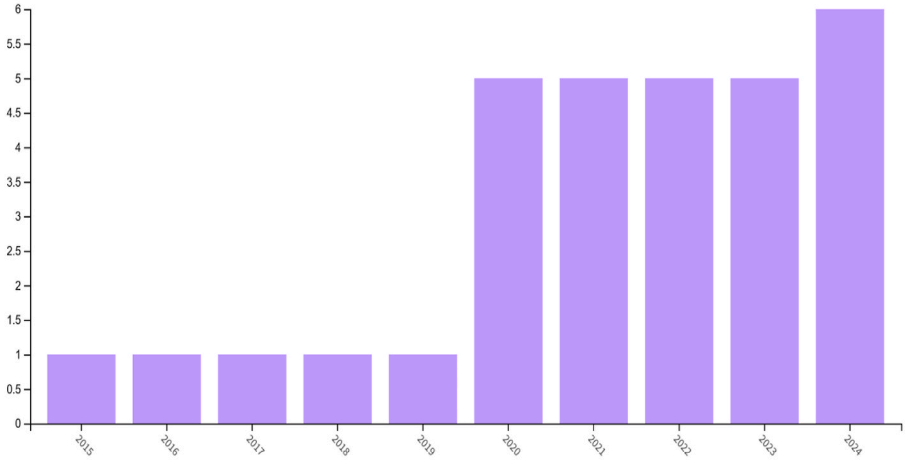
(c) Keyword search: Energy load forecasting (96 documents)

Fig. 15. Frequency of publications for keywords related to energy prediction (2012–2024).

Fig. 15 illustrates the annual frequency of publications for each keyword related to energy prediction. Panel A demonstrates the growing interest in energy consumption forecasting, Panel B highlights consistent research activity in energy demand prediction, and Panel C reflects the increasing focus on energy load forecasting in recent years.To further refine the literature search, the representative keywords from both categories were combined using the Boolean operators "AND and OR. This approach aimed to identify studies that explore different types of ensemble or hybrid models applied to the calculation of energy consumption. By using OR, synonyms and closely related terms within the same category were grouped. The AND operator was then used to intersect the two categories, linking machine learning models with energy prediction, thus narrowing the results to studies that directly address the objectives of this review. Figure 16 shows the results of the total number of publications in Scopus and Web of Science, which were found with the search equations cited in Table 9.



(a) Search Query: Scopus



(b) Search Query: Web of Science

Fig. 16. Annual distribution of publications from Scopus/Web of Science using the final search equation 2012–2024.

The final result of this combination of words and operators can be seen in Table 9, which summarizes the different search equations in the two selected databases.

Table 9
Search queries and the number of selected articles.

Database	Search Query	Time Range	Initial Results	Selected Articles
Scopus	TITLE-ABS-KEY ("ensemble model" OR "hybrid model") AND TITLE-ABS-KEY ("energy consumption forecasting" OR "energy demand prediction" OR "energy load forecasting")	2012–2024	65	40
Web of Science	TS=("ensemble model" OR "hybrid model") AND TS=("energy consumption forecasting" OR "energy demand prediction" OR "energy load forecasting")	2012–2024	42	10
Other		2012–2024	17	17

As can be seen the combination of articles between Web of Science, Scopus and other sources was a total of 124, with some articles duplicated in the databases. Scopus provides the majority of results with 65 articles, while Web of Science returns 42 and 17 articles were selected from different databases (Table 9). A final selection of 67 articles has been analyzed.

To facilitate the understanding of the models and algorithms discussed throughout the manuscript, a comprehensive list of abbreviations used in this review is provided below (Table 10)

Table 10
List of Model Abbreviations Used in This Review

Abbreviation	Full Term/Meaning
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
CNN-LSTM	Convolutional Neural Network–Long Short-Term Memory
BiGRU	Bidirectional Gated Recurrent Unit
GRU	Gated Recurrent Unit
PLSTM	Parallel Long Short-Term Memory
EDA-LSTM	Enhanced Dual Attention Long Short-Term Memory
Transformer	Attention-Based Neural Network
GRU-MMattention	Gated Recurrent Unit with Multi-Modal Attention
Improved-CNN	Enhanced Convolutional Neural Network
SSA	Singular Spectrum Analysis
VMD	Variational Mode Decomposition
EMD	Empirical Mode Decomposition
QES	Quadratic Exponential Smoothing
ANFIS	Adaptive Neuro-Fuzzy Inference System
GEP	Gene Expression Programming
ADE	Adaptive Differential Evolution
PSO	Particle Swarm Optimization
SVM	Support Vector Machine
LSSVR	Least Squares Support Vector Regression
RBFNN	Radial Basis Function Neural Network
GRNN	General Regression Neural Network
MLR	Multiple Linear Regression
MLP	Multilayer Perceptron
DNN	Deep Neural Network
RNN	Recurrent Neural Network
BP	Backpropagation (Neural Network)
ELM	Extreme Learning Machine
ESN	Echo State Network
GMDH	Group Method of Data Handling
GP	Genetic Programming
FCRBM	Factorized Conditional Restricted Boltzmann Machine
GWDO	Genetic Wind-Driven Optimization
SOS	Symbiotic Organism Search
JS	Jellyfish Search
SARIMA	Seasonal Autoregressive Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
GM(1,1)	Grey Model (First order, one variable)
XGBoost	Extreme Gradient Boosting
AdaBoost	Adaptive Boosting
LightGBM	Light Gradient Boosting Machine
RF	Random Forest
RFR	Random Forest Regressor
ETR	Extra Trees Regressor
VR	Voting Regressor
SVC	Support Vector Classification
SGD	Stochastic Gradient Descent
KNN	K-Nearest Neighbors
K-means	K-means Clustering Algorithm
K-means++	Improved K-means Clustering
GAN	Generative Adversarial Network
AFF-PROP	Affinity Propagation
DL	Deep Learning
ML	Machine Learning
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning

Data availability

No data was used for the research described in the article.

References

[1] Hong T, Fan S. Probabilistic electric load forecasting: a tutorial review. Int J Forecast Jul. 2016;32(3):914–38. https://doi.org/10.1016/j.ijforecast.2015.11.011.

[2] Montuori L, Alcazar-Ortega M, Vargas-Salgado C, Manso-Borras I. Enabling the natural gas system as smart infrastructure: metering technologies for customer applications. In: 2020 global Congress on electrical engineering (GC-ElecEng). IEEE; Sep. 2020. p. 96–100. https://doi.org/10.23919/GC-ElecEng48342.2020.9286291.

[3] Montuori L, Alcazar-Ortega M, Manso-Borras I, Vargas-Salgado C. Communication technologies in smart grids of natural gas: new challenges. In: 2020 global Congress on electrical engineering (GC-ElecEng). IEEE; Sep. 2020. p. 101–5. https://doi.org/10.23919/GC-ElecEng48342.2020.9286281.

[4] Montuori L, Alcázar-Ortega M. District heating as demand response aggregator: estimation of the flexible potential in the Italian peninsula. Energies Oct. 2021;14(21):7052. https://doi.org/10.3390/en14217052.

[5] Hatziargyriou N, Asano H, Iravani R, Marnay C. Microgrids. IEEE Power Energy Mag Jul. 2007;5(4):78–94. https://doi.org/10.1109/MPAE.2007.376583.

[6] Slowik M, Urban W. Machine learning short-term energy consumption forecasting for microgrids in a manufacturing plant. Energies May 2022;15(9):3382. https://doi.org/10.3390/en15093382.

[7] Aslam S, Herodotou H, Mohsin SM, Javaid N, Ashraf N, Aslam S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. Renew Sustain Energy Rev Jul. 2021;144:110992. https://doi.org/10.1016/j.rser.2021.110992.

[8] Zhang G, Eddy Patuwo B, Hu MY. Forecasting with artificial neural networks. Int J Forecast Mar. 1998;14(1):35–62. https://doi.org/10.1016/S0169-2070(97)00044-7.

[9] Breiman L. Random forests. Mach Learn 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.

[10] Vargas-Salgado C, Borras IM, Alcazar-Ortega M, Montuori L. Floating PV solar-powered systems: design applied to an irrigators community in the Valencia region – spain. In: 2020 global Congress on electrical engineering (GC-ElecEng). IEEE; Sep. 2020. p. 88–95. https://doi.org/10.23919/GC-ElecEng48342.2020.9286279.

[11] Zhang D, Han X, Deng C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE Journal of Power and Energy Systems Sep. 2018;4(3):362–70. https://doi.org/10.17775/CSEEJPES.2018.00520.

[12] Li S, Oshnoei A, Blaabjerg F, Anvari-Moghaddam A. Hierarchical control for microgrids: a survey on classical and machine learning-based methods. Sustainability Jun. 2023;15(11):8952. https://doi.org/10.3390/su15118952.

[13] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: a machine learning case study. Energy Build Jun. 2012;49:591–603. https://doi.org/10.1016/j.enbuild.2012.03.010.

[14] Xiao L, Wang J, Yang X, Xiao L. A hybrid model based on data preprocessing for electrical power forecasting. Int J Electr Power Energy Syst Jan. 2015;64:311–27. https://doi.org/10.1016/j.ijepes.2014.07.029.

[15] Fan G-F, Han Y-Y, Li J-W, Peng L-L, Yeh Y-H, Hong W-C. A hybrid model for deep learning short-term power load forecasting based on feature extraction statistics techniques. Expert Syst Appl Mar. 2024;238:122012. https://doi.org/10.1016/j.eswa.2023.122012.

[16] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat Oct. 2001;29(5). https://doi.org/10.1214/aos/1013203451.

[17] Singh T, Sethi AS, Singh H, Rana PS, Kumar N, Obaidat MS. Two-tier ensemble model for demand side prediction in smart grid environment. In: 2019 IEEE global communications conference (GLOBECOM). IEEE; Dec. 2019. p. 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9013536.

[18] Yan K, Yang F, Zhou X, Liang W. Parallel LSTM enhanced forecasting system for dependable and secure household energy consumption. In: 2021 IEEE 23rd int conf on high performance computing & communications; 7th int conf on data Science & Systems; 19th int conf on Smart City; 7th int conf on dependability in sensor, Cloud & big data Systems & application (HPCC/DSS/SmartCity/DependSys). IEEE; Dec. 2021. p. 1988–95. https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00297.

[19] Doan VK, Doan DN, Nguyen BH. Improving energy demand prediction use deep learning network. In: 2024 9th international conference on applying new technology in green buildings (ATiGB). IEEE; Aug. 2024. p. 323–7. https://doi.org/10.1109/ATiGB63471.2024.10717679.

[20] Ou Ali IH, Agga A, Ouassaid M, Maaroufi M, Elrashidi A, Kotb H. Predicting short-term energy usage in a smart home using hybrid deep learning models. Front Energy Res 2024;12(Sep). https://doi.org/10.3389/fenrg.2024.1323357.

[21] Niu D, Yu M, Sun L, Gao T, Wang K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. Appl Energy May 2022;313:118801. https://doi.org/10.1016/j.apenergy.2022.118801.

[22] Peng J, Kimmig A, Wang J, Liu X, Niu Z, Ovtcharova J. Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. Energy Build Oct. 2021;249:111211. https://doi.org/10.1016/j.enbuild.2021.111211.

[23] Bakare MS, Abdulkarim A, Shuaibu AN, Muhamad MM. A hybrid long-term industrial electrical load forecasting model using optimized ANFIS with gene expression programming. Energy Rep Jun. 2024;11:5831–44. https://doi.org/10.1016/j.egyr.2024.05.045.

[24] Sunder R, et al. An advanced hybrid deep learning model for accurate energy load prediction in smart building. Energy Explor Exploit Nov. 2024;42(6):2241–69. https://doi.org/10.1177/01445987241267822.

[25] Jrhilifa I, Ouadi H, Jilbab A, Mounir N. Forecasting smart home electricity consumption using VMD-Bi-GRU. Energy Effic Apr. 2024;17(4):35. https://doi.org/10.1007/s12053-024-10205-0.

[26] Yan K, Yang F, Zhou X, Liang W. Parallel LSTM enhanced forecasting system for dependable and secure household energy consumption. In: 2021 IEEE 23rd int conf on high performance computing & communications; 7th int conf on data Science & Systems; 19th int conf on Smart City; 7th int conf on dependability in sensor, Cloud & big data Systems & application (HPCC/DSS/SmartCity/DependSys). IEEE; Dec. 2021. p. 1988–95. https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00297.

[27] Liang S, Deng T, Huang A, Liu N, Jiang X. Energy consumption prediction using the GRU-MMattention-LightGBM model with features of Prophet decomposition. PLoS One Jan. 2023;18(1):e0277085. https://doi.org/10.1371/journal.pone.0277085.

[28] Arslan S. A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data. PeerJ Comput Sci Jun. 2022; 8:e1001. https://doi.org/10.7717/peerj-cs.1001.

[29] Taheri S, Talebjedi B, Laukkanen T. Electricity demand time series forecasting based on empirical mode decomposition and long short-term memory. Energy Eng J Assoc Energy Eng 2021;118(6):1577–94. https://doi.org/10.32604/EE.2021.017795.

[30] Sharma S, Agrawal RK, Tripathi MM. Synergism of recurrent neural network and fuzzy logic for short term energy load forecasting. In: 2020 fourth international conference on computing methodologies and communication (ICCMC). IEEE; Mar. 2020. p. 165–9. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00033.

[31] Somu N, M R GR, Ramamritham K. A hybrid model for building energy consumption forecasting using long short term memory networks. Appl Energy Mar. 2020;261:114131. https://doi.org/10.1016/j.apenergy.2019.114131.

[32] Liu Z, Wang X, Zhang Q, Huang C. Empirical mode decomposition based hybrid ensemble model for electrical energy consumption forecasting of the cement grinding process. Measurement May 2019;138:314–24. https://doi.org/10.1016/j.measurement.2019.02.062.

[33] Zeng Y-R, Zeng Y, Choi B, Wang L. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy May 2017; 127:381–96. https://doi.org/10.1016/j.energy.2017.03.094.

[34] Pombo DV, Bindner HW, Spataru SV, Sørensen PE, Bacher P. Increasing the accuracy of hourly multi-output solar power forecast with physics-informed machine learning. Sensors Jan. 2022;22(3):749. https://doi.org/10.3390/s22030749.

[35] Cao Y, Liu G, Sun J, Bavirisetti DP, Xiao G. PSO-stacking improved ensemble model for campus building energy consumption forecasting based on priority feature selection. J Build Eng Aug. 2023;72:106589. https://doi.org/10.1016/j.jobe.2023.106589.

[36] Chou J, Truong D. Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning. Int J Energy Res Mar. 2021;45(3):4581–612. https://doi.org/10.1002/er.6125.

[37] Xu L, et al. Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm. Energy May 2021;222:119955. https://doi.org/10.1016/j.energy.2021.119955.

[38] Hafeez G, et al. A novel accurate and fast converging deep learning-based model for electrical energy consumption forecasting in a smart grid. Energies May 2020; 13(9):2244. https://doi.org/10.3390/en13092244.

[39] Xu L, Hou L, Zhu Z, Li Y, Lei T. A hybrid method based on SVM integrated improved PSO Algorithm for electrical energy consumption forecasting of crude oil pipeline. In: Operations, monitoring, and maintenance; materials and joining, vol. 3. American Society of Mechanical Engineers; Sep. 2020. https://doi.org/10.1115/IPC2020-9434.

[40] Zhao H, Guo S. Annual energy consumption forecasting based on PSOCA-GRNN model. Abstr Appl Anal 2014;2014:1–11. https://doi.org/10.1155/2014/217630.

[41] Tran D-H, Luong D-L, Chou J-S. Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings. Energy Jan. 2020;191: 116552. https://doi.org/10.1016/j.energy.2019.116552.

[42] Liao Z, Su M, Ning G, Liu Y, Wang T, Zhou J. A novel stacked generalization ensemble-based hybrid PSVM-PMLP-MLR model for energy consumption prediction of copper foil electrolytic preparation. IEEE Access 2021;9:5821–31. https://doi.org/10.1109/ACCESS.2020.3048714.

[43] Mubarak H, Sanjari MJ, Stegen S, Abdellatif A. Improved active and reactive energy forecasting using a stacking ensemble approach: steel industry case Study. Energies Oct. 2023;16(21):7252. https://doi.org/10.3390/en16217252.

[44] Sahraei MA, Çodur MK. Prediction of transportation energy demand by novel hybrid meta-heuristic ANN. Energy Jun. 2022;249:123735. https://doi.org/10.1016/J.ENERGY.2022.123735.

[45] Gupta G, Mathur S, Mathur J, Nayak BK. Blending of energy benchmarks models for residential buildings. Energy Build Aug. 2023;292:113195. https://doi.org/10.1016/j.enbuild.2023.113195.

[46] Khan A-N, Iqbal N, Rizwan A, Ahmad R, Kim D-H. An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings. Energies May 2021;14(11):3020. https://doi.org/10.3390/en14113020.

[47] Wu D, Hur K, Xiao Z. A GAN-enhanced ensemble model for energy consumption forecasting in large commercial buildings. IEEE Access 2021;9:158820–30. https://doi.org/10.1109/ACCESS.2021.3131185.

[48] Singh T, Sethi AS, Singh H, Rana PS, Kumar N, Obaidat MS. Two-Tier ensemble model for demand side prediction in smart Grid environment. In: 2019 IEEE global communications conference (GLOBECOM). IEEE; Dec. 2019. p. 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9013536.

[49] Swide S, Marhoon A. Energy demand prediction based on deep learning techniques. Iraqi Journal for Electrical and Electronic Engineering Dec. 2023;19 (2):83–9. https://doi.org/10.37917/ijeee.19.2.10.

[50] Khan PW, Byun Y-C. Genetic algorithm based optimized feature engineering and hybrid machine learning for effective energy consumption prediction. IEEE Access 2020;8:196274–86. https://doi.org/10.1109/ACCESS.2020.3034101.

[51] Wang X, Meng M. A hybrid neural network and ARIMA model for energy consumption forcasting. J Comput May 2012;7(5). https://doi.org/10.4304/jcp.7.5.1184-1190.

[52] Hachache R, et al. Energy load forecasting techniques in smart grids: a cross-country comparative analysis. Energies May 2024;17(10):2251. https://doi.org/10.3390/en17102251.

[53] Padilha G, et al. Transformer-based hybrid forecasting model for multivariate renewable energy. Appl Sci Oct. 2022;12(21):10985. https://doi.org/10.3390/APP122110985. 2022, Vol. 12, Page 10985.

[54] Lyes A, Saoud S, Almarzouqi H, Hussein R. Cascaded deep hybrid models for multistep household energy consumption forecasting [Online]. Available: https://arxiv.org/pdf/2207.02589. [Accessed 26 April 2025].

[55] Dogra A, Anand A, Bedi J. Consumers profiling based federated learning approach for energy load forecasting. Sustain Cities Soc Nov. 2023;98:104815. https://doi.org/10.1016/j.scs.2023.104815.

[56] de Mattos Neto PSG, et al. Energy consumption forecasting for smart meters using extreme learning machine ensemble. Sensors Dec. 2021;21(23):8096. https://doi.org/10.3390/s21238096.

[57] Helli SS, Tanberk S, Demir O. Forecasting energy consumption using deep learning in smart cities. In: 2022 international conference on artificial intelligence of things

(ICAIoT). IEEE; Dec. 2022. p. 1–6. https://doi.org/10.1109/ICAIoT57170.2022.10121846.

[58] Waqas Khan P, Byun Y-C, Lee S-J, Park N. Machine learning based hybrid system for imputation and efficient energy demand forecasting. Energies May 2020;13 (11):2681. https://doi.org/10.3390/en13112681.

[59] Xiao J, Li Y, Xie L, Liu D, Huang J. A hybrid model based on selective ensemble for energy consumption forecasting in China. Energy Sep. 2018;159:534–46. https://doi.org/10.1016/j.energy.2018.06.161.

[60] Yuan C, Liu S, Fang Z. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. Energy Apr. 2016;100:384–90. https://doi.org/10.1016/j.energy.2016.02.001.

[61] Chou J-S, Tran D-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy Dec. 2018;165:709–26. https://doi.org/10.1016/j.energy.2018.09.144.

[62] Coelho VN, et al. A self-adaptive evolutionary fuzzy model for load forecasting problems on smart grid environment. Appl Energy May 2016;169:567–84. https://doi.org/10.1016/j.apenergy.2016.02.045.

[63] Talaat M, Elkholy MH, Alblawi A, Said T. Artificial intelligence applications for microgrids integration and management of hybrid renewable energy sources. Artif Intell Rev Sep. 2023;56(9):10557–611. https://doi.org/10.1007/s10462-023-10410-w.

[64] Parol M, Piotrowski P, Kapler P, Piotrowski M. Forecasting of 10-Second power demand of highly variable loads for Microgrid operation control. Energies Feb. 2021;14(5):1290. https://doi.org/10.3390/en14051290.

[65] Shirzadi N, Nasiri F, El-Bayeh C, Eicker U. Optimal dispatching of renewable energy-based urban microgrids using a deep learning approach for electrical load and wind power forecasting. Int J Energy Res Mar. 2022;46(3):3173–88. https://doi.org/10.1002/er.7374.

[66] Nahid FA, Ongsakul W, Singh JG, Roy J. Short-term customer-centric electric load forecasting for low carbon microgrids using a hybrid model. Energy Systems Oct. 2024. https://doi.org/10.1007/s12667-024-00704-5.

[67] Thanh PN, Cho M-Y, Chang C-L, Chen M-J. Short-term three-phase load prediction with advanced metering infrastructure data in Smart solar microgrid based convolution neural network bidirectional gated recurrent unit. IEEE Access 2022;10:68686–99. https://doi.org/10.1109/ACCESS.2022.3185747.